DRUG DISCOVERY AGAINST MALARIA PARASITE: DRUG REPOSITIONING STRATEGY

BY

JOO HWAN NO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Eric Oldfield, Chair
Professor Robert B. Gennis
Associate Professor Anne M. Baranger
Assistant Professor Douglas A. Mitchell

Abstract

Malaria, caused by *Plasmodium* spp., causes ~1 million deaths each year and there are ever present problems due to drug resistance. In this work, I used drug-repositioning strategy with bisphosphonates to discover a lead that is active against malaria *in vivo*. The developmental procedures include, *in vitro* high throughput screening, x-ray crystallography and other biophysical techniques. First, a computational method was developed to elucidate the target of bisphosphonate in malaria parasite and, second, in-house synthesized library of prenyl synthase inhibitors was used to find the lead, an analog of zoledronate, against the parasite, which the x-ray structure bound to target enzyme, farnesyl diphosphate synthase (FPPS), was solved. Also, the effect of lipophilic bisphosphonate against liver stage malaria parasite was investigated. Lastly, I studied the mechanism and inhibition of IspH, the last enzyme of nonmevalonate pathway, for discovery of a novel target for infectious disease.

Acknowledgements

First of all, I thank my research advisor, Professor Eric Oldfield for his patient guidance and mentoring throughout my study. I also thank other members in my committee, Professor Robert B. Gennis, Professor Anne M. Baranger and Professor Douglas Mitchell.

I would like to show appreciations to Dr. Dushyant Mukkamala, Dr. Fu-Yang Lin, Dr. Yonghui Zhang, Dr. Rong Cao, Dr. Ke Wang, Dr. Michael Hudock, Dr. Sujoy Mukherjee, Dr Annette Leon-Rossell, Yi-Liang Liu, Francisco Guerra, Kilannin Krysiak, Kyle Bergan, and other current Oldfield group members.

I also thank Cindy Dodds, coordinator of the Center for Biophysics and Computational Biology, for her help and guidance during my years in the Program.

Finally, I am forever indebted to my mom, Hee Sung Park, for her love of sacrifice and dedication to me, and to my father, Kyoung Tai No, for raising me with scientific inspiration as well as showing me how a warm-hearted scientist could be. Also I thank my brother, Jin Hwan No, for his hidden support and love, and Jihye Kwak for going through hard times with me during my years in Urbana. Thank you God, how great you are.

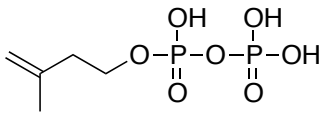
Table of Contents

Chapter 1

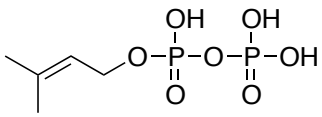
Isoprenoid Biosynthesis Pathway as Drug Target

1.1 Isoprenoid Pathway

Isoprenoids are the largest class of natural products consisted by more than 50,000 distinct small molecules(1), and among numerous enzymes that are involved in isoprenoid biosynthesis are validated drug targets. One of the most well known example is statin (Lipitor), which targets the early step of cholesterol biosynthesis. Bisphosphonates, such as Fosamax, are used to treat bone resorption disease and targets the enzyme farnesyl diphosphate synthase (FPPS) in the middle stage of isoprenoid biosynthesis, while anti-infectives such as Lamisil targets downstream enzyme of sterol biosynthesis in yeasts and fungi. The initial step of isoprenoid pathway is the formation of isopentenyl diphosphate (**1**, IPP) and dimethylallyl diphosphate (**2**, DMAPP). Pathogenic bacteria and protozoan such as *Mycobacterium tuberculosis* and *Plasmodium falciparum* uses the Rohmer or nonmevalonate pathway to produce IPP and DMAPP(2), but incase of humans and bacteria such as *Staphylococcus aureus*, they are formed by the mevalonate pathway(3)(Figure 1.1). In the nonmevalonate pathway, the last two enzymes are IspG (GcpE) and IspH (LytB), containing Fe_4S_4 clusters(4, 5) to carry out reduction reactions from 2-*C*-methyl-D-erythritol-2,4-cyclo-diphosphate (MEcPP) to *E*-1-hydroxyl-2-methyl-but-2-enyl 4-diphosphate (HMBPP) and HMBPP to IPP and DMAPP (shown in Figure 1.1).



1



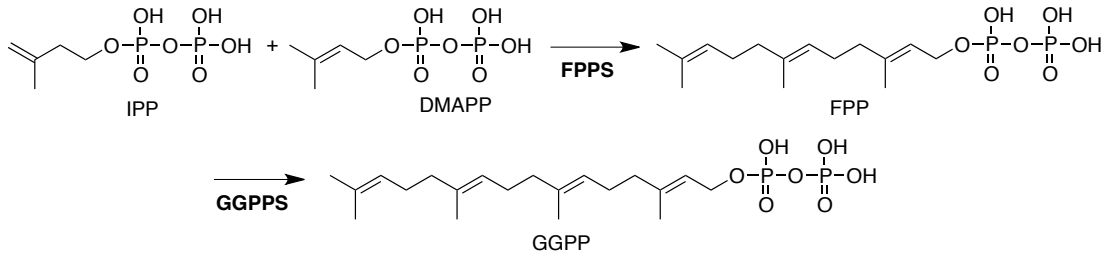
2

After IPP and DMAPP are formed, condensation of two products via a "head-to-tail" mechanism produces geranyl diphosphate (GPP) and farnesyl diphosphate (FPP) by FPPS and further reaction with IPP catalyzed by the enzyme geranylgeranyl diphosphate synthase (GGPPS) yields the C_{20} species, geranylgeranyl diphosphate (GGPP)(6, 7), shown in Figure 1.1. After FPP and GGPP are formed, they are used in protein prenylation, known to be important for cell survival and signaling pathway. FPP and GGPP can also condensed to form squalene via presqualene diphosphate(8) or phytoene via prephytoene diphosphate.

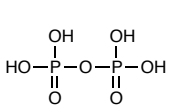
In many organisms, squalene is converted to squalene epoxide by squalene epoxidase (SE) which then cyclized to form lanosterol and amiodrone targets the site as inhibition of action, shown in Figure 1.1. After couple of additional steps, lanosterol is converted to cholesterol in humans and ergosterol or episterol in yeasts, fungi, and protozoa. In *S. aureus*, dehydrosqualene, *cis* form of squalene, is converted to a carotenoid pigment, staphyloxanthin(9), a virulence factor responsible for protecting the organism from reactive oxygen species from its host. In Figure 1.1, several additional enzymes in the isoprenoid pathway that are known to be important targets of well-known inhibitors are depicted. By elucidating their structures, mechanism of action and inhibition, designing of novel inhibitors can take place in knowledge-based fashion.

1.2 FPPS and GGPPS: Structure, Mechanism and Inhibition

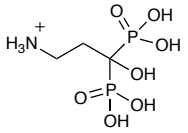
Through the nonmevalonate or mevalonate pathway, IPP and DMAPP are produced and further condensed by FPPS and GGPPS to form FPPS and GGPPS, as shown in scheme below.



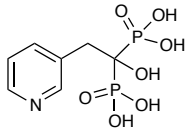
In the pathway, the bisphosphonate class of drugs targets FPPS to treat bone resorption disease, but for over decades the target enzyme and their mechanism of action were unknown. The interest from our group in the system was drawn from couple of unexpected observations. First, with collaborations with Urbina and Docampo(10), we discovered that *Trypanosoma cruzi* contained very high levels of condensed phosphate, such as diphosphate, shown below. Such discovery led us to explore nonhydrolysable PPi analogs such as pamidronate and risedronate (Actonel) could possibly inhibit parasite cell growth. In fact, this turns out to be the case(10, 11), but the enzyme target for these inhibitors were still unknown. The second observation was that



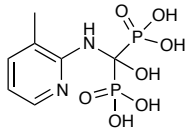
diphosphate



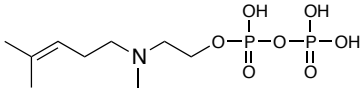
Pamidronate



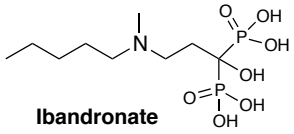
Actonel



3



4



Ibandronate

nitrogen-containing analogs of GPP such as **4**, was found to be a potent inhibitor of terpene cyclases and the overall structure resembles that of ibandronate. This observation suggests that cationic bisphosphonates could act as carbocation/diphosphate isosteres, blocking isoprenoid biosynthesis(12). The last of initial observation was that compound such as **3** was a herbicide developed by Zeneca(13) and was found to be a low nanomolar inhibitor of a daffodil FPPS(13). Since **3** was previously known to be active against bone resorption(14), we concluded that the bone-resorption(15) drugs would act by mimicking a carbocation reaction intermediate. The target for such compounds were soon confirmed crystallographically by Hosfield et al(16). Fallowing work in our group showed that pamidronate is active against cutaneous leishmaniasis(17), Figure 1.2, by inhibiting FPPS which lead to blocking ergosterol biosynthesis(11), opening up the new era of clinical use of bisphosphonates as anti-infectives(18).

Further discovery of bisphosphonates showed killing of tumor cells(19) and activation of γδ T cells(20), which also targets to kill tumor cell(21). As a result, the interest on killing tumor cells resulted in small clinical trials on pamidronate(22) and zoledronate(23). Recent large scale clinical study of 1803 patients with breast cancer showed3 30% of reduction of recurrence of disease in patients treated with postsurgery plus zoledronate(24). Such drug-repositioning strategies are now under rigorous investigation in various fields of disease for further development and protozoa including malaria is now one of disease of interest.

1.3 LytB (IspH): a Novel Target Containing Fe_4S_4-Cluster

The nonmevalonate pathway is attractive site for drug development, since the pathway is not present in human but is essential for survival of pathogen, including *Plasmodium falciparum* and *Mycobacterium tuberculosis*, and currently fosmidomycin, targeting DXP reductoisomerase in the pathway (Figure 1.1), is under Phase III clinical trial. In the pathway, the last reaction of the pathway is carried out by the enzyme IspH(25, 26). Our collaborations with Jomaa and Ermler, we reported(27) that the overall structure of the enzyme is a unique, trefoil-like fold with Fe_3S_4 cluster positioned at the center of the protein (Figure 1.3A, B) and a similar structure was again reported by Grawert et al(28). Both the reported structures contained 3Fe and not 4Fe, incoherent with the spectrum of electron paramagnetic resonance (EPR), chemical analysis, and activity results, which all pointed to an Fe_4S_4 cluster, so we carried out computational modeling approach to construct an Fe_4S_4 model and performed docking of HMBPP substrate to unique fourth Fe in oxidized IspH, via its 1-OH group(27), Figure 1.3C.

Interestingly, recent X-ray crystallographic structure of LytB(29) shows that HMBPP does actually bind to 4Fe cluster via O-1 as proposed in our group, and the binding mode of HMBPP

to Fe_4S_4 generated from docking(27) is almost identical to that determined by X-ray crystallography, Figure 1.3D, with lignad RMSD of 0.3Å. But the catalytic mechanism of Fe4S4 with $2H^+/2e^-$ reduction, the removal of the 1-OH oxygen to form the IPP and DMAPP, is not elucidated. Based on our crystallographic and bioinformatics investigation, we proposed(27) that E126 it the critical residue in catalysis, providing the H^+ required for activity. The essential nature of E126 was then observed by other groups(28), and also by creating E126A mutant we though to "trap" the reaction intermediate which could be informative for drawing out the enzyme involved in catalytic mechanism. So we carried out EPR and electron-nuclear double resonance (ENDOR) spectroscopy(30). With extensive investigation on the LytB, we were able to understand in details on the catalytic mechanism, which could lead us to develop novel inhibitors against the enzyme that are known to be essential for deleterious pathogens.

1.4 Figures

Figure 1.1. Drug targets and their inhibitors in the isoprenoid pathway

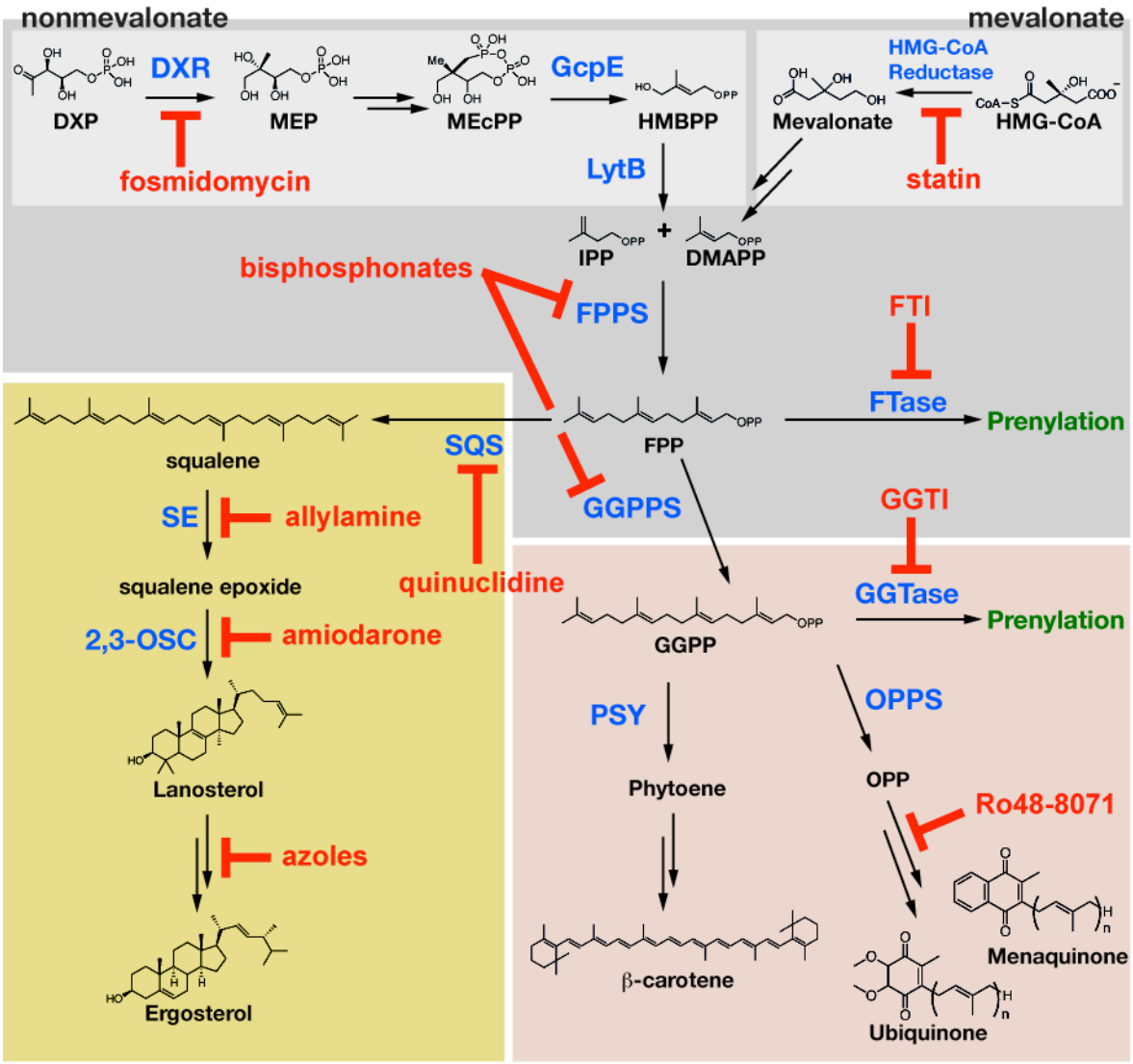


Figure 1.2. Effect of pamidronate on cutaneous leishmaniasis in mice (**A**) dose dependent inhibition on lesion progression; (**B**) cure of infection in treated mouse

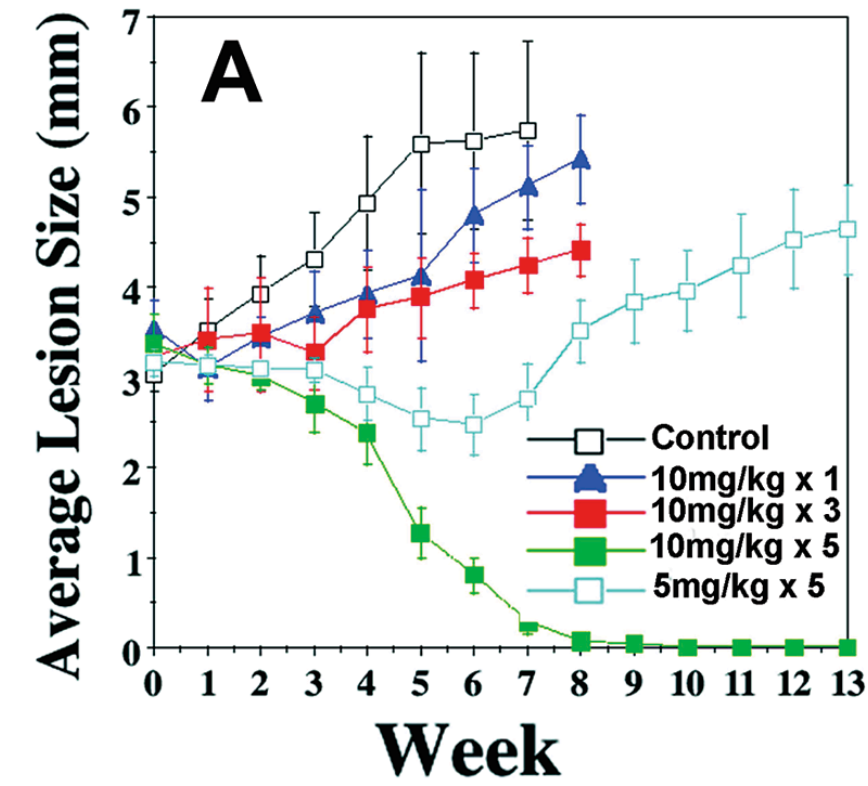
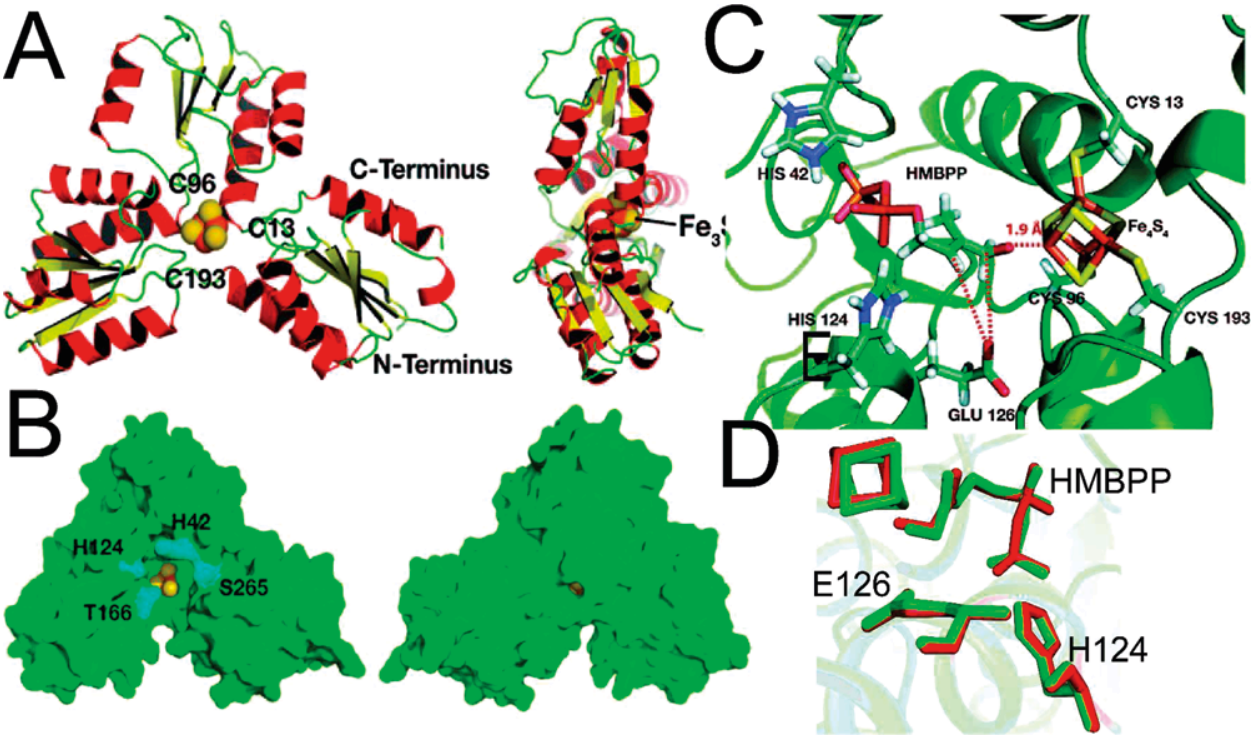


Figure 1.3. Structural results for IspH (LytB): (**A**,**B**) crystal structure results for *Aquifex aeolicus* IspH; (**C**) initial docking pose for HMBPP to oxidized IspH Fe_4S_4 cluster obtained by using the "open-form" structure; (**D**) Comparison of HMBPP bound to IspH from X-ray (green) and docking (red).



1.5 References

1. Poulter CD (2009) Bioorganic chemistry. A natural reunion of the physical and life sciences. *J Org Chem* 74(7):2631-2645 (in eng).
2. Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat Prod Rep* 16(5):565-574 (in eng).
3. Goldstein JL & Brown MS (1990) Regulation of the mevalonate pathway. *Nature* 343(6257):425-430 (in eng).
4. Seemann M*, et al.* (2002) Isoprenoid biosynthesis through the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (GcpE) is a [4Fe-4S] protein. *Angew Chem Int Ed Engl* 41(22):4337-4339 (in eng).
5. Wolff M*, et al.* (2003) Isoprenoid biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB/IspH) from *Escherichia coli* is a [4Fe-4S] protein. *FEBS Lett* 541(1-3):115-120 (in eng).
6. Thulasiram HV & Poulter CD (2006) Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity. *J Am Chem Soc* 128(49):15819-15823 (in eng).
7. Ohnuma S, Hirooka K, Ohto C, & Nishino T (1997) Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase. Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. *J Biol Chem* 272(8):5192-5198 (in eng).
8. Epstein WW & Rilling HC (1970) Studies on the mechanism of squalene biosynthesis. The structure of presqualene pyrophosphate. *J Biol Chem* 245(18):4597-4605 (in eng).

9. Pelz A*, et al.* (2005) Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. *J Biol Chem* 280(37):32493-32498 (in eng).

10. Urbina JA*, et al.* (1999) *Trypanosoma cruzi* contains major pyrophosphate stores, and its growth *in vitro* and *in vivo* is blocked by pyrophosphate analogs. *J Biol Chem* 274(47):33609-33615 (in eng).

11. Martin MB*, et al.* (2001) Bisphosphonates inhibit the growth of *Trypanosoma brucei*, *Trypanosoma cruzi, Leishmania donovani, Toxoplasma gondii,* and *Plasmodium falciparum*: a potential route to chemotherapy. *J Med Chem* 44(6):909-916 (in eng).

12. Luckman SP*, et al.* (1998) Nitrogen-containing bisphosphonates inhibit the mevalonate pathway and prevent post-translational prenylation of GTP-binding proteins, including Ras. *J Bone Miner Res* 13(4):581-589 (in eng).

13. Cromartie TH, Fisher KJ, & Grossman JN (1999) The discovery of a novel site of action for herbicidal bisphosphonates. (Translated from English) *Pestic Biochem Phys* 63(2):114-126 (in English).

14. Sunberg RJ, Ebetino FH, Mosher CT, & Roof CF (1991) Designing Drugs for Stronger Bones. (Translated from English) *Chemtech* 21(5):304-309 (in English).

15. Martin MB, Arnold W, Heath HT, 3rd, Urbina JA, & Oldfield E (1999) Nitrogen-containing bisphosphonates as carbocation transition state analogs for isoprenoid biosynthesis. *Biochem Biophys Res Commun* 263(3):754-758 (in eng).

16. Hosfield DJ*, et al.* (2004) Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. (Translated from English) *Journal of Biological Chemistry* 279(10):8526-8529 (in English).

17. Rodriguez N, *et al.* (2002) Radical cure of experimental cutaneous leishmaniasis by the bisphosphonate pamidronate. (Translated from English) *J Infect Dis* 186(1):138-140 (in English).

18. Paterson R (2002) Pamidronate next on list as potential cure for leishmaniasis. (Translated from English) *Lancet Infect Dis* 2(9):515-515 (in English).

19. Derenne S, *et al.* (1999) Zoledronate is a potent inhibitor of myeloma cell growth and secretion of IL-6 and MMP-1 by the tumoral environment. *J Bone Miner Res* 14(12):2048-2056 (in eng).

20. Kunzmann V, Bauer E, & Wilhelm M (1999) Gamma/delta T-cell stimulation by pamidronate. *N Engl J Med* 340(9):737-738 (in eng).

21. Kunzmann V, *et al.* (2000) Stimulation of gammadelta T cells by aminobisphosphonates and induction of antiplasma cell activity in multiple myeloma. *Blood* 96(2):384-392 (in eng).

22. Wilhelm M, *et al.* (2003) Gammadelta T cells for immune therapy of patients with lymphoid malignancies. *Blood* 102(1):200-206 (in eng).

23. Dieli F, *et al.* (2007) Targeting human {gamma}delta} T cells with zoledronate and interleukin-2 for immunotherapy of hormone-refractory prostate cancer. *Cancer Res* 67(15):7450-7457 (in eng).

24. Gnant M, *et al.* (2009) Endocrine therapy plus zoledronic acid in premenopausal breast cancer. *N Engl J Med* 360(7):679-691 (in eng).

25. Rohrich RC, *et al.* (2005) Reconstitution of an apicoplast-localised electron transfer pathway involved in the isoprenoid biosynthesis of *Plasmodium falciparum*. *FEBS Lett* 579(28):6433-6438 (in eng).

26. Eberl M, *et al.* (2003) Microbial isoprenoid biosynthesis and human gammadelta T cell activation. *FEBS Lett* 544(1-3):4-10 (in eng).

27. Rekittke I, *et al.* (2008) Structure of (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase, the terminal enzyme of the non-mevalonate pathway. *J Am Chem Soc* 130(51):17206-17207 (in eng).

28. Grawert T, *et al.* (2009) Structure of active IspH enzyme from Escherichia coli provides mechanistic insights into substrate reduction. *Angew Chem Int Ed Engl* 48(31):5756-5759 (in eng).

29. Grawert T, *et al.* (2010) Probing the reaction mechanism of IspH protein by x-ray structure analysis. *Proc Natl Acad Sci U S A* 107(3):1077-1081 (in eng).

30. Wang W, *et al.* (2010) Bioorganometallic mechanism of action, and inhibition, of IspH. *Proc Natl Acad Sci U S A* 107(10):4522-4527 (in eng).

Chapter 2

Bisphosphonate Inhibition of a *Plasmodium* Farnesyl Diphosphate Synthase and a General Method for Predicting Cell-Based Activity from Enzyme Data

2.1 Notes and Acknowledgements

We thank Raymond Hui of the Structural Genomics Consortium (Toronto, Canada) for providing the *P. vivax* protein expression system. This work was supported by the United States Public Health Service (NIH grant GM65307).

2.2 Introduction

Malaria causes ~10^6 deaths annually, so there is considerable interest in developing novel drugs to treat this disease. One pathway of interest is that involved in the isoprenylation of signaling proteins, where protein-farnesyl transferase inhibitors have shown promise(1), since they act by preventing the post-translational modification of proteins such as Ras. A second potential target involves inhibition of the enzymes (such as farnesyl diphosphate synthase, FPPS) that produce the isoprene diphosphates used in protein prenylation, and in recent work we showed that bisphosphonates, drugs that target FPPS and are widely used in treating bone resorption diseases(2), had both *in vitro* and *in vivo* activity against *Plasmodium* parasites(3, 4). There were, however, some puzzling aspects to the results(4), in particular, the most active species against *P. falciparum* had little activity against a variety of expressed FPPS enzymes, while other, known potent FPPS inhibitors, had low activity against the parasite.

There are three likely explanations for these observations. First, it might be that FPPS is not the actual (or only) target. Second, there might be surprising differences in the sensitivity of *Plasmodium* FPPS and human (or other) FPPS enzymes, since *Plasmodium* FPPS can also produce geranylgeranyl diphosphate (GGPP), due to a smaller barrier to chain elongation at the end of the binding site. Third, it might just be difficult to obtain good correlations between enzyme and cell inhibition data, due to neglect of "transport" issues. This latter point is an exceptionally important one(5) and is of very broad general interest for drug development, since while enzyme inhibition assays can generally be carried out very rapidly (and accurately), cell based assays are more expensive and take much longer. Indeed, in some cases, cell data may be essentially uncorrelated with enzyme inhibition. A good recent example of this is that of the inhibition of undecaprenyl diphosphate synthase from *Streptococcus pneumoniae* reported by Peukert et al.(6), in which an $R^2 = 0.03$ is found from the reported cell/enzyme pIC_{50} (= $-\log_{10}$ IC_{50}) results, obviously complicating lead optimization using enzyme based assays.

In this work, we report the inhibition of a *Plasmodium* FPPS (from *P. vivax*) by a library of 26 bisphosphonates whose activity against *P. falciparum* cell growth *in vitro* was reported previously(4). The *P. falciparum* and *P. vivax* enzymes have very similar sequences (73% identity, 89% similarity) and catalytic site residues, but we find that there is essentially no correlation between the cell and enzyme pIC_{50} values ($R^2 = 0.06$). We find, however, that regression of cell, enzyme, and two other descriptors leads to a remarkable improvement in the correlation between experimental and predicted cell pIC_{50} (to $R^2 = 0.74$), suggesting that small descriptor sets might be used to predict cell activity results in many other systems, nine of which (including other anti-parasitic, anti-bacterial, anti-viral, anti-diabetic and anti-cancer drug leads) are investigated in this work. Overall, our results show that cell activity in 10 very varied

systems can be well predicted by using enzyme inhibition data, when combined with the combinatorial descriptor search approach.

2.3 Results and Discussion

We show in Figure 2.1 the structures of the 26 bisphosphonates (**1**-**26**) investigated previously(4). These compounds have IC_{50} values in *P. falciparum* growth inhibition varying from 1.43 to 222 μM(4) with the most active species also having *in vivo* activity in a *P. berghei* ANKA suppressive test (up to an 80% reduction in parasitemia). The IC_{50}/pIC_{50} values in FPPS inhibition are presented in Table 2.1, together with the previously published cell growth inhibition results(4). When the cell pIC_{50} and enzyme pIC_{50} values are compared, we see that there is essentially no correlation between the two data sets (R^2 = 0.06), Figure 2.2a, resulting in a leave-2–out set of predictions having an R^2=0.01. For example, the most active species in FPPS inhibition is **15** (IC_{50} = 400 nM), but this compound has poor activity *in vitro* (33.6 μM, Table 2.1). It might therefore be thought that FPPS is not the actual (or only) target for these bisphosphonates, or that there are small but significant differences in structure between the *P. falciparum* and *P. vivax* enzymes. However, on further inspection of Figure 2.2a, there appear to be several possible "clusters" of compounds (**A**, **B**, **C** in Figure 2.2a), with the more active species in cells having more hydrophobic features. For example, in cluster **A**, known bisphosphonates such as risedronate (**20**), ibandronate (**16**), zoledronate (**23**) and minodronate (**17**) all have very poor (>70μM) cell activity (Table 2.1), even though they have good activity in the enzyme assay. Since these compounds are among the most hydrophilic ones investigated (with an average SlogP = -5.00, for 9 compounds in this cluster), it is possible that their poor cell based activity is due to poor transport. This idea receives support from the observation that a

second class of bisphosphonates (such as **13**, **15**; cluster **B**, with average SlogP = -4.60, for 5 compounds) that have improved hydrophobicity also exhibit improved cell activity, and compounds which have the highest cell-based activity (**1**-**6**, Figure 2.1, Cluster **C**) also contain the most hydrophobic side chains (average SlogP = -3.60, for 10 compounds). But how can we put these qualitative observations on a more quantitative footing? Can we predict cell activity from the enzyme inhibition results? And if this turns out to be possible in the *Plasmodium* system: can the method be generalized to other systems, something that would be of very broad, general interest.

The results shown in Table 2.1 and Figure 2.2a clearly indicate that there is essentially no correlation between the cell pIC_{50} and enzyme pIC_{50} values. Plus, the range in cell activity (~100x) is much larger than the range in enzyme inhibition (~5x range in IC_{50}). Why is this? While, as noted above, there are several possibilities (other targets, poor *P. vivax*/*P. falciparum* similarity), the observation that many of the bisphosphonates investigated have similar enzyme inhibition activity at the same time that their cell-based activity covers a ~100x range, together with the observation that cell based activity appears to increase as hydrophobicity increases, suggests the possibility that cell permeability may be particularly important in governing overall cell activity. To test this hypothesis, we chose to represent the cell based activity (pIC_{50} (cell)) mathematically as:

$$pIC_{50}(\text{cell}) = a \bullet pIC_{50}(\text{enzyme}) + b\ \text{SlogP} + c \qquad (1)$$

where SlogP is a computed oil/water partition coefficient. We chose Slog P (rather than say clogP, etc.) without any particular basis, a point we return to shortly. Data were analyzed by using linear regression, from which we find a = 1.26, b = 0.62 and c = -0.34. Now, when the pIC_{50} (cell, experiment) results are correlated with the predicted pIC_{50} (cell) results, we find a

good correlation between experiment and prediction, with an $R^2 = 0.66$. Given this promising result, we then sought to find other descriptors and descriptor combinations that might give improved predictivity, rather than restricting ourselves to SlogP. In some initial calculations we used enzyme pIC_{50} values together with 3 descriptors, or 4 random descriptors chosen from MOE(7), that gave the best experimental versus predicted R^2 values. However, this approach was found to lead to over-fitting, since some predictions using randomized cell activity data had R^2 values of ca. 0.5. We thus eventually chose the following equation:

$$pIC_5 \text{ (cell)} = a \bullet pIC_{50} \text{ (enzyme)} + b \bullet B + c \bullet C + d \qquad (2)$$

in which two additional descriptors (B,C) were chosen from a combinatorial search, performed on a large set of descriptors (Table 2.4) in MOE(7). We chose to fix pIC_{50} (enzyme) as one descriptor, since the objective is to predict cell activity from enzyme activity, where standard QSAR and structure-based methods can be employed to optimize enzyme inhibition. There are 230 descriptors available in MOE, and a full search and cross-validation of all (26,335) two-descriptor combinations for a dataset containing 26 compounds is lengthy (~30 hours). However, we found that it was not necessary to use all the 230 descriptors, for two reasons: First, because some descriptors are linear combinations of other descriptors in the database. And second, because some descriptors are Boolean (with predominantly either zeros or ones) and do not contribute significantly in the linear regression. In the set of 26 *P. vivax* FPPS inhibitors, after eliminating redundant and Boolean descriptors, we obtained 150 descriptors, of which all combinations of two (B1, C1; B2, C2;……; B150, C150) were investigated, the coefficients (a…d) being determined via linear regression. The 10 descriptor combinations (pIC_{50}; B_i; C_j) giving the highest R^2 values (for the training set experimental *versus* predicted pIC_{50} correlation) are shown in Table 2.

As can be seen in Table 2.2, the best experimental/predicted pIC_{50} correlation has an R^2 = 0.77 (Figure 2.2b), a major increase over the cell/enzyme correlation (R^2 = 0.06) alone. For the top 10 predictions, 9/10 contain a SlogP or logP(o/w) term, indicating the importance of hydrophobicity in cell based activity. And in 6/10 cases, this SlogP or logP(o/w) term has the largest relative contribution to the correlation (Table 2.2). In the four cases where this term does not dominate, enzyme activity dominates, but as can be seen in Table 2.2, both enzyme pIC_{50} and logP-like terms are important in all cases, and the R^2 range is very small, 0.74 to 0.77. So, both enzyme and "logP" contribute to cell activity, with the cell activity results apparently being well predicted based on the enzyme inhibition and two-descriptor model. Of course, considerable care needs to be taken in order to not over-fit data when using this approach. We tested this in several ways: First, we used a leave-two-out (L2O) method in which we systematically left out all possible pairs of data points ($^{26}C_2$=26x25/2 = 325 calculations), re-evaluating the a-d coefficients for all possible B, C descriptors (total time ~12 hours). The results of this L2O test set of predictions are shown graphically in Figure 2.2c, where we find an R^2 = 0.74 for the experimental/predicted pIC_{50} test set correlation, close to the R^2 = 0.77 for the training set (Figure 2.2b). Second, we scrambled the cell pIC_{50} values and repeated the L2O test set prediction, 10 times. On average, the R^2 value was 0.16 for this scrambled test set, strongly supporting the idea that the approach, when applied to real cell growth inhibition data, is highly predictive. Using larger numbers of descriptors increased the R^2 value for both the unscrambled and scrambled L2O data sets, and is not recommended, at least for relatively small data sets.

These results encouraged us to see to what extent this approach might enable us to improve cell activity predictions, based on enzyme activity and the combinatorial descriptor search, for a variety of other systems. We first chose to investigate two other systems we

reported on previously: *Dictyostelium discoideum*(8) (a eukaryote, used as a screen for bone resorption drugs that target FPPS), and *Leishmania donovani*(3, 9), a causative agent of visceral leishmaniasis. Both systems (or very close relatives) have been shown by knockouts or overexpression to require FPPS, and that FPPS is the target for bisphosphonate drugs or inhibitors.

For *Dictyostelium discoideum* growth inhibition(8) the cell/enzyme R^2 value is 0.49 (Figure 2.3a) or $R^2 = 0.46$ for a L2O enzyme only test set, and this increases to $R^2 = 0.70$ for the training set, using two additional descriptors (Figure 2.3b; Table 2.5), with pIC_{50} (enzyme) being the most important contributor to the correlation. This is not unexpected since, unlike the *Plasmodium* results, we do see a correlation between cell and enzyme data, Figure 3a. The leave-two-out test set results (Table 2.3, Figure 2.3c) are likewise good ($R^2 = 0.70$) and as with the *Plasmodium* results, the scrambled data set has essentially no predictivity ($R^2 = 0.07$). Slightly improved results are obtained for *Leishmania donovani* cell growth inhibition. Here, the raw cell/enzyme correlation is $R^2 = 0.55$ (Table 2.3, Figure 2.3d), translating to a L2O enzyme-based test set $R^2 = 0.47$. The experimental /predicted cell pIC_{50} correlation increases to $R^2 = 0.87$ for the training set using two additional descriptors (Figure 2.3e and Table 2.6) and 0.80 for the L2O test set results (Figure 2.3f), while using scrambled cell data yields $R^2 = 0.11$ (Table 2.3). We should also note here that in both *D. discoideum* and *L. donovani*, FPPS inhibition results for the specific organisms were not available, so we used a composite data set of *Trypanosoma brucei*(10), *T. cruzi*(11), *L. major*(12) and human(13) FPPS results to model the *D. discoideum* and *L. donovani* data, just as we used *P. vivax* data to interpret the *P. falciparum* results. Naturally, it is reasonable to believe that using e.g. *L. donovani* FPPS to model *L. donovani* growth inhibition will be preferred to using e.g. *L. major* FPPS inhibition data. However, the

predictions are still good, with for *P. falciparum*, *D. discoideum* and *L. donovani*, the average R^2 value increasing from 0.37 (cell/enzyme L2O test set predictions) to 0.72, using pIC_{50} (enzyme) plus two descriptors, while the average R^2 for the scrambled cell activity data sets is $R^2 = 0.11$. But, given that all of the inhibitors studied here are bisphosphonates and the targets protozoa or a unicellular eukaryote, it is fair to ask: Is this method robust? Does it apply to other types of cell and inhibitor? Is it, in short, a general method for probing cell activity based on enzyme inhibition?

Anti-bacterial and Anti-Viral Systems.

To test the generality of this approach, we investigated two bacterial and two viral systems. In each case, the enzyme targeted in the enzyme assay was the same as that targeted in the cell based assay. In one recent study, workers at Novartis(6) reported the discovery of a series of novel inhibitors (tetramic, tetronic acids and dihydropyridin-2-ones) of the enzyme undecaprenyl diphosphate synthase (UPPS) that also had good cellular activity(6). These inhibitors are of interest since UPPS is a potentially important target for anti-infective development, because undecaprenyl diphosphate is used in Lipid A biosynthesis, and UPPS inhibition is expected to block bacterial cell wall biosynthesis. Plus, the structures of some UPPS enzymes are now known(14, 15). *Streptococcus pneumoniae* enzyme and (approximate) cell growth inhibition results (inhibition at 1 μM, 2 μM, 4 μM etc. inhibitor concentrations) were reported, but as with our *Plasmodium* results, there is little correlation between the cell growth and enzyme inhibition data sets ($R^2 = 0.03$; Figure 2.3g), even though in this case the enzyme target (*S. pneumoniae* UPPS) is thought to be exactly the same as that present in the bacterium, *Streptococcus pneumoniae*(6). The small range in cell activity (128x) combined with the limited precision of the cell data (the IC_{50} values were rounded off by factors of 2) makes this a challenging case.

Even so, using the reported enzyme pIC_{50} results together with two descriptors yields a considerable improvement, with R^2 increasing from 0.10 for the L2O test set using enzyme data only, to $R^2 = 0.69$ (Figure 2.3i) for the L2O test set predictions using two extra descriptors, with R^2 for the L2O test set using scrambled data being 0.14, Table 2.3. As can be seen in Table 2.7, descriptors such as SlogP and logS (also found in the *Plasmodium* data sets) are again found here and, when combined with enzyme inhibition data, enable greatly improved predictions of the cell activity results. Another set of anti-bacterials are glutamic acid analogs that inhibit *S. pneumoniae* glutamate racemase (MurI)(16). In this case, there is a moderate correlation between cell growth and MurI inhibition ($R^2 = 0.61$), and this increases to 0.72 (training) or 0.68 (leave-two-out) test set predictions using the combinatorial descriptor search, Table 2.3 and Figure 2.4 and Table 2.8, with the R^2 for scrambled cell activity test set data being 0.13.

In addition to these results with anti-bacterial systems, we find good predictivity in an anti-viral assay using hepatitis C virus protease inhibitors(17). Here, there is again a very poor correlation between enzyme and cell pIC_{50} values (obtained in a cell-based replicon assay) with $R^2 = 0.03$ (Figure 2.4d). This increases to $R^2 = 0.78$ (training) and $R^2 = 0.77$ (leave-two-out) test set results (Figures 2.4e and 2.4f) with incorporation of two additional descriptors. In this case, the best results were obtained by using PAMPA (parallel artificial membrane permeability assay) as one of the descriptors (determined experimentally by Li et. al.(17)). In these calculations, we added reported PAMPA, Caco-2, log P-neutral, log D-pH 7.4, clog P, and mlog P values to the 230 potential descriptor fields in MOE, with PAMPA being selected in the descriptor search as the most important descriptor, followed by the enzyme pK_i (Table 2.9). So, the predictivity of the method can be improved when experimental data on permeability is available.

We also investigated a second series of anti-virals, novel phthalimide-analog inhibitors of HIV-1 integrase(18). The results of enzyme pIC_{50}, HIV-1 replication in a cell-based assay (pEC_{50}), as well as cell-based toxicity (pCC_{50}) have been reported(18). There is only a very poor correlation between the enzyme pIC_{50} and cell-based replication activity (pEC_{50}), with $R^2 = 0.12$ (Figure 2.4g; $R^2 = 0.01$ for the L2O enzyme based test set). Using the combinatorial descriptor search, we find that pCC_{50} is the most important descriptor, enabling an experimental *versus* predicted pIC_{50} R^2=0.69 (Figure 2.4h). That is, pCC_{50} makes the major contribution to the apparent cell (HIV) replication results (Table 2.10), an observation of interest for further optimization of this class of compounds. And on average, for the four anti-bacterial/antiviral systems, we see that R^2 increases from 0.22 to 0.70 (L2O test sets) on addition of two descriptors, about the same as that found for the protozoal/primitive eukaryote systems with scrambled cell data, $R^2 = 0.13$, about the same as the $R^2 = 0.11$ found with the protozoa/simple eukaryotes.

Mammalian Cell Lines: Anti-Diabetic and Anti-cancer Drugs.

Finally, we wanted to test the method described above on three mammalian cell lines, in diabetes and cancer assays, as opposed to the primarily anti-infective drugs/targets described in the preceding sections. We investigated three systems: acyl urea inhibitors of human liver glycogen phosphorylase (hlGP)(19) having activity in rat hepatocytes; 1,4-dihydroindeno[1,2-c]pyrazoles targeting a KDR kinase in NIH3T3 cells(20), and an indazole-pyridine series of protein kinase B/Akt inhibitors, active in pancreatic cancer cells(21).

In the case of the acyl urea inhibitors of human liver glycogen phosphorylase, the enzyme/cell correlation was already evident (R^2= 0.55), and this improved to R^2=0.74 for the L2O test set of 35 compounds, Table 2.3 and Figures 2.5a-c and Table 2.11. The scrambled data

was, as expected, not predictive, with an R^2=0.19. Results for the KDR kinase (Table 2.3, Figures 2.5d-f and Table 2.12) were slightly worse (L2O test set R^2=0.58), while results with the Akt kinase (Table 2.3, Figures 2.5g-i) were good, with an R^2=0.68 for the test set (Table 2.13).

2.4 Conclusions

The results described above are of interest for several reasons. First, we screened a library of bisphosphonates against a *P. vivax* FPPS enzyme. The results were surprising since we found that inhibition of the *P. vivax* enzyme was essentially uncorrelated with growth inhibition by these same bisphosphonates of the malaria parasite, *P. falciparum*. However, when pIC_{50} results together with two additional descriptors were regressed against cell growth inhibition pIC_{50} results, we found good predictivity of the cell results, with an R^2 = 0.77 (0.74 for a leave-two-out test set). This observation then led to a general method for predicting the cell-based activity of a compound from its activity in an enzyme assay, in which we use a combinatorial descriptor search to choose two descriptors (B,C) that, in combination with the enzyme pIC_{50} results, enabled relatively good cell activity predictions. For the ten diverse systems investigated here, which cover anti-parasitic, anti-bacterial, anti-viral, anti-diabetes and anti-cancer drug candidates, the average R^2 value increases from 0.28 for the leave-two-out test set predictions using solely enzyme inhibition data, to 0.70 for the leave two-out test sets with two additional descriptors, to be compared with 0.13 for predictions based on randomized cell activity data. This approach should be particularly useful in optimizing cell activity when target structures are known, since structure-based (x-ray, QSAR) methods can be used to optimize enzyme inhibition, with the combinatorial descriptor search then being used to make good predictions of cell based activity. Plus, when experimental data on permeability is available, this can be included in the

descriptor field as another potential descriptor, to be used in cell activity predictions. A larger number of descriptors than the (on average) 124 we employed can also of course be used, at the expense of computational time, which can become large (~300 hours for a 230 descriptor based series of calculations, including validation tests). The approach described above should thus enable the use of both crystallographic and enzyme QSAR data for enzyme inhibition optimization, with cell growth inhibition (or viral replication) activity being described using the combinatorial descriptor approach. The method should also be applicable in some cases to modeling toxicity and *in vivo* results as well, when sufficient data is available.

2.5 Materials and Methods

P. vivax FPPS expression and inhibition.

A clone encoding *P. vivax* FPPS (PlasmoDB gene ID: Pv092040) with an N-terminally His_6-tagged fusion protein and a TEV protease site was expressed in *E. coli* BL21-codon Plus (DE3) RIL (Stratagene) at 20 °C in baffled flasks. Cells were lysed in the presence of lysozyme (Sigma), Benzonase® Nuclease (Novagen), and a protease inhibitor cocktail (Sigma), and protein purified chromatographically by using a Ni-NTA resin. The *P. vivax* FPPS assays were carried out by using 96-well plates with 150 μL reaction mixture in each well. The condensation of geranyl diphosphate (GPP) with isopentenyl diphosphate (IPP) was monitored at room temperature by using a continuous spectrophotometric assay for phosphate releasing enzymes in a reaction mixture containing 25 mM Hepes and 2.5 mM $MgCl_2$ at pH 7.4. The inhibitors were preincubated with the enzyme for 30 min at room temperature. The IC_{50} values were obtained from fitting the dose-response curve using Prism 4.0(22).

Computational Aspects.

We performed a complete combinatorial descriptor search in MATLAB(23) (running on a Dual 2 GHz PowerPC G5 or a Dual-Core Intel Xeon Mac Pro) , using linear regression of enzyme pIC_{50} plus all possible two-descriptor combinations ($^{150}C_2$=11175 combinations for a database containing 150 descriptors, for example) against the cell activities, pIC_{50}(cell). For the leave-two-out (L2O) cross-validation, the same exhaustive search was performed, using a training set obtained by leaving out two compounds from the initial data set. This process was repeated for all pair-wise combinations of the compounds (e.g. for 26 compounds, 1,2; 1,3;…1,26; 2,3; 2,4; ….; 24,25; 24,26; 25,26). So, if there are 26 compounds (as in *P. vivax*), then each compound is left out (and predicted) 25 times. In this way, for *P. vivax* FPPS we obtained 25 predicted values for each compound in the dataset, and these were then averaged to give the overall leave-two-out predicted value, for each compound. To test whether the correlations obtained might occur by chance, we randomly scrambled the cell activities, then performed the leave-two-out validation on the scrambled data. The process was repeated 10 times and the mean R^2 values are reported. For the 10 systems investigated, each scrambling analysis requires ~10-90 hours, depending upon the number of descriptors (80-150) and the number of compounds (21-102) in the data set. Linearly dependent descriptors were eliminated based on the occurrence of rank deficiency in the descriptor correlation matrix (in MATLAB[23]).

2.6 Tables and Figures

Table 2.1. Experimental IC_{50} and pIC_{50} values for *P.falciparum* growth inhibition and *P. vivax* FPPS inhibition.

Compound[a]	IC_{50} (Enzyme, μM)	pIC_{50} (Enzyme)[b]	IC_{50} (Cell, μM)	pIC_{50} (Cell)[c]
1	0.63	6.20	1.43	5.84
2	4.09	5.39	2.12	5.67
3	0.61	6.21	2.9	5.54
4	0.71	6.15	3.32	5.48
5	0.68	6.17	3.82	5.42
6	0.63	6.20	4.34	5.36
7	0.63	6.20	6.43	5.19
8	0.91	6.04	10.6	4.97
9	0.69	6.16	11.1	4.95
10	0.49	6.31	11.5	4.94
11	0.88	6.06	22.1	4.66
12	2.32	5.63	22.6	4.65
13	0.45	6.35	22.8	4.64
14	0.82	6.09	28.2	4.55
15	0.4	6.40	33.6	4.47
16	0.92	6.04	73.8	4.13
17	0.64	6.19	90.2	4.04
18	5.18	5.29	112	3.95
19	0.71	6.15	115	3.94
20	0.59	6.23	129	3.89
21	0.78	6.11	131	3.88
22	1.24	5.91	159	3.80
23	0.79	6.10	167	3.78
24	0.87	6.06	184	3.74
25	98.62	4.01	200	3.70
26	0.9	6.05	222	3.65

[a] See Figure 2.1 for chemical structures

[b] The enzyme assays were performed on FPPS from *Plasmodium vivax*

[c] The cell assay data is for *Plasmodium falciparum(4)*

Table 2.2. Top ten "enzyme plus 2-descriptor" combinations with coefficients and relative contributions for *P. falciparum* growth inhibition predictions.

Rank	R^2	Coefficient a[1] (enzyme)	Relative Contribution (enzyme)	Coefficient b[1]	Descriptor B[2]	Relative Contribution (B)	Coefficient c[1]	Descriptor C[2]	Relative Contribution (C)	Constant,
1	0.77	1.0407	0.72	0.52375	logP(o/w)	1.00	-0.0028993	E_sol	0.47	-3.2113
2	0.76	1.3081	0.84	-0.5824	logS	1.00	4.6753	E_oop	0.42	-3.7232
3	0.76	1.0806	0.69	0.56454	logP(o/w)	1.00	4.9535	GCUT_SMR_0	0.45	1.4987
4	0.75	1.3317	1.00	0.42892	SlogP	0.75	0.011375	PEOE_VSA_NEG	0.50	-3.7753
5	0.75	0.98902	0.70	0.51324	logP(o/w)	1.00	-292.05	BCUT_PEOE_1	0.44	-203.38
6	0.75	1.2344	0.85	0.62848	SlogP	1.00	1.0862	E_strain	0.30	-0.23834
7	0.75	1.342	1.00	0.5162	SlogP	0.89	0.17337	chi1_C	0.40	-1.8939
8	0.75	1.2085	1.00	0.49972	SlogP	0.96	1.8143	GCUT_SLOGP_1	0.42	0.49546
9	0.75	1.044	0.74	0.51263	logP(o/w)	1.00	0.86654	BCUT_SLOGP_0	0.44	1.5301
10	0.74	1.0733	1.00	0.44935	SlogP	0.97	0.0098483	PEOE_VSA-1	0.49	-0.61185

[1] from the equation: pIC_{50} (cell) = a•pIC_{50} (enzyme) + b•B + c•C + d

[2] Descriptors selected are described in Table 2.4

Table 2.3. Results for ten different systems showing the correlations between enzyme and cell activities, and the correlations between experimental and predicted cell activities.

Target	Cell	R^2, Enzyme vs. Cell	R^2, Enzyme Test Set[a]	R^2, Training Set[b]	R^2, Test Set[a]	R^2, Scrambled[c]	Descriptors tested	Number of compounds
FPPS	*Plasmodium vivax*	0.06	0.01	0.77	**0.74**	0.16	150	26
FPPS	*Dictyostelium discoideum*	0.49	0.46	0.70	**0.70**	0.07	123	102
FPPS	*Leishmania donovani*	0.55	0.47	0.87	**0.8**	0.11	138	21
UPPS	*Streptococcus pneumoniae*	0.03	0.10	0.71	**0.69**	0.14	109	27
MurI	*Streptococcus pneumoniae*	0.61	0.57	0.72	**0.68**	0.13	123	42
HCV NS3 Protease	Hu-7	0.03	0.02	0.78	**0.77**	0.12	142	34
HIV-1 integrase	MT4	0.12	0.04	0.69	**0.64**	0.13	126	30
HLG Phosphorylase	Rat hepatocytes	0.55	0.48	0.71	**0.74**	0.19	132	35
KDR kinase	NIH 3T3	0.36	0.30	0.60	**0.58**	0.09	117	42
Akt Kinase	MiaPaCa-2 (pancreatic cancer)	0.41	0.35	0.68	**0.68**	0.12	80	48
Average		0.32	0.28	0.72	**0.70**	0.13	124	42

[a] Correlation obtained from a leave-two-out analysis
[b] Correlations based on cell activity predictions using a combinatorial descriptor search
[c] Average correlation obtained from a leave-two-out analysis of 10 sets of scrambled data

Table 2.4. List of descriptors used in the Combinatorial Descriptor Search[a]

Code	Descripton	Number of Occurences[b]
log(PAMPA)	Log of parallel artificial membrane permeability	10
log(Caco-2)	Log of permeability coefficient for transport into Caco-2 cells	1
log P-neutral	Log of the concentrations of un-ionized compound between water and octanol	
log D-pH 7.4	Log of the ratio of the sum of concentrations of solute's various forms between water and octanol	
clogP	Log P calculated by Sybyl 7.3(24)	
mlogP	Measured log P calculated by Sybyl 7.3(24)	
diameter	Largest value in the distance matrix	
petitjean	Value of (diameter - radius) / diameter.	
petitjeanSC	Petitjean graph Shape Coefficient(25): (diameter - radius) / radius.	
VDistEq	If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of $\log_2 m - p_i \log_2 p_i / m$ where p_i is the number of distance matrix entries equal to i.	
VDistMa	If m is the sum of the distance matrix entries then VDistMa is defined to be the sum of $\log_2 m - D_{ij} \log_2 D_{ij} / m$ over all i and j.	
weinerPath	Wiener path number: half the sum of all the distance matrix entries(26, 27)	
weinerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3(26)	
BCUT_PEOE_0 BCUT_PEOE_1 BCUT_PEOE_2 BCUT_PEOE_3	The BCUT descriptors(28) are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(b_{ij}) where b_{ij} is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.	1
BCUT_SLOGP_0 BCUT_SLOGP_1 BCUT_SLOGP_2 BCUT_SLOGP_3	The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge(29).	1
BCUT_SMR_0 BCUT_SMR_1 BCUT_SMR_2 BCUT_SMR_3	The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge(29).	
GCUT_PEOE_0 GCUT_PEOE_1 GCUT_PEOE_2 GCUT_PEOE_3	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.	

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
GCUT_SLOGP_0 GCUT_SLOGP_1 GCUT_SLOGP_2 GCUT_SLOGP_3	The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge(29).	3
GCUT_SMR_0 GCUT_SMR_1 GCUT_SMR_2 GCUT_SMR_3	The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge(29).	1
a_count	Number of atoms (including implicit hydrogens). This is calculated as the sum of $(1 + h_i)$ over all non-trivial atoms i.	
a_IC	Atom information content (total). This is calculated to be a_ICM times n.	
a_ICM	Atom information content (mean). This is the entropy of the element distribution in the molecule (including implicit hydrogens but not lone pair pseudo-atoms). Let n_i be the number of occurrences of atomic number i in the molecule. Let $p_i = n_i / n$ where n is the sum of the n_i. The value of a_ICM is the negative of the sum over all i of $p_i \log p_i$.	
b_1rotN	Number of rotatable single bonds. Conjugated single bonds are not included (e.g., ester and peptide bonds).	
b_1rotR	Fraction of rotatable single bonds: b_1rotN divided by b_heavy.	1
b_count	Number of bonds (including implicit hydrogens). This is calculated as the sum of $(d_i/2 + h_i)$ over all non-trivial atoms i.	
b_rotN	Number of rotatable bonds. A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors.	
b_rotR	Fraction of rotatable bonds: b_rotN divided by b_heavy.	1
b_single	Number of single bonds (including implicit hydrogens). Aromatic bonds are not considered to be single bonds.	
chi0v	Atomic valence connectivity index (order 0)(30, 31). This is calculated as the sum of $1/\mathrm{sqrt}(v_i)$ over all heavy atoms i with $v_i > 0$.	
chi0v_C	Carbon valence connectivity index (order 0). This is calculated as the sum of $1/\mathrm{sqrt}(v_i)$ over all carbon atoms i with $v_i > 0$.	
chi1v	Atomic valence connectivity index (order 1)(30, 31). This is calculated as the sum of $1/\mathrm{sqrt}(v_i v_j)$ over all bonds between heavy atoms i and j where $i < j$.	
chi1v_C	Carbon valence connectivity index (order 1). This is calculated as the sum of $1/\mathrm{sqrt}(v_i v_j)$ over all bonds between carbon atoms i and j where $i < j$.	

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
Weight	Molecular weight (including implicit hydrogens) with atomic weights taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	
b_heavy	Number of bonds between heavy atoms.	
a_nC	Number of carbon atoms: #{Z_i | Z_i = 6}.	
chi0	Atomic connectivity index (order 0)(30, 31). This is calculated as the sum of 1/sqrt(d_i) over all heavy atoms i with d_i > 0.	
chi0_C	Carbon connectivity index (order 0). This is calculated as the sum of 1/sqrt(d_i) over all carbon atoms i with d_i > 0.	
chi1	Atomic connectivity index (order 1)(30, 31). This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between heavy atoms i and j where i < j.	
chi1_C	Carbon connectivity index (order 1). This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between carbon atoms i and j where i < j.	1
VAdjEq	Vertex adjacency information (equality): $-(1-f)\log_2(1-f) - f\log_2 f$ where $f = (n^2 - m) / n^2$, n is the number of heavy atoms and m is the number of heavy-heavy bonds. If f is not in the open interval (0,1), then 0 is returned.	
VAdjMa	Vertex adjacency information (magnitude): $1 + \log_2 m$ where m is the number of heavy-heavy bonds. If m is zero, then zero is returned.	
zagreb	Zagreb index: the sum of d_i^2 over all heavy atoms i.	
balabanJ	Balaban's connectivity topological index(32).	
PC+ Q_PC+ PEOE_PC+	Total positive partial charge: the sum of the positive q_i. Q_PC+ is identical to PC+ which has been retained for compatibility.	2
PC- Q_PC- PEOE_PC-	Total negative partial charge: the sum of the negative q_i. Q_PC- is identical to PC- which has been retained for compatibility.	
RPC+ Q_RPC+ PEOE_RPC+	Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i. Q_RPC+ is identical to RPC+ which has been retained for compatibility.	1
RPC- Q_RPC- PEOE_RPC-	Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i. Q_RPC- is identical to RPC- which has been retained for compatibility.	5
PEOE_VSA+0	Sum of v_i where q_i is in the range [0.00,0.05).	2
PEOE_VSA+1	Sum of v_i where q_i is in the range [0.05,0.10).	
PEOE_VSA-0	Sum of v_i where q_i is in the range [-0.05,0.00).	
PEOE_VSA-1	Sum of v_i where q_i is in the range [-0.10,-0.05).	1

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
Q_VSA_FHYD PEOE_VSA_FHYD	Fractional hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FNEG PEOE_VSA_FNEG	Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FPNEG PEOE_VSA_FPNEG	Fractional negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	1
Q_VSA_FPOL PEOE_VSA_FPOL	Fractional polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	4
Q_VSA_FPOS PEOE_VSA_FPOS	Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FPPOS PEOE_VSA_FPPOS	Fractional positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	
Q_VSA_HYD PEOE_VSA_HYD	Total hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2. The v_i are calculated using a connection table approximation.	2
Q_VSA_NEG PEOE_VSA_NEG	Total negative van der Waals surface area. This is the sum of the v_i such that q_i is negative. The v_i are calculated using a connection table approximation.	1
Q_VSA_PNEG PEOE_VSA_PNEG	Total negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2. The v_i are calculated using a connection table approximation.	
Q_VSA_POL PEOE_VSA_POL	Total polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2. The v_i are calculated using a connection table approximation.	1
Q_VSA_POS PEOE_VSA_POS	Total positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative. The v_i are calculated using a connection table approximation.	
Q_VSA_PPOS	Total positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2. The v_i are calculated using a connection table approximation.	1

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
lip_acc	The number of O and N atoms.	
opr_nrot	The number of rotatable bonds(33).	
E	Value of the potential energy. The state of all term enable flags will be honored (in addition to the term weights). This means that the current potential setup accurately reflects what will be calculated.	1
E_ang	Angle bend potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
E_ele	Electrostatic component of the potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	2
E_nb	Value of the potential energy with all bonded terms disabled. The state of the non-bonded term enable flags will be honored (in addition to the term weights).	3
E_oop	Out-of-plane potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	5
E_sol	Solvation energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	1
E_stb	Bond stretch-bend cross-term potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
E_str	Bond stretch potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	1
E_strain	Local strain energy: the current energy minus the value of the energy at a near local minimum. The current energy is calculated as for the E descriptor. The local minimum energy is the value of the E descriptor after first performing an energy minimization. Current chirality is preserved and charges are left undisturbed during minimization. The structure in the database is not modified (results of the minimization are discarded).	1
E_tor	Torsion (proper and improper) potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	3
E_vdw	van der Waals component of the potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
Kier1	First kappa shape index: $(n\text{-}1)^2 / m^2$ (31).	
Kier2	Second kappa shape index: $(n\text{-}1)^2 / m^2$ (31).	1
Kier3	Third kappa shape index: $(n\text{-}1)\ (n\text{-}3)^2 / p_3^2$ for odd n, and $(n\text{-}3)\ (n\text{-}2)^2 / p_3^2$ for even n (31).	1

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
KierA1	First alpha modified shape index: $s\,(s\text{-}1)^2 / m^2$ where $s = n + a$ (31).	2
KierA2	Second alpha modified shape index: $s\,(s\text{-}1)^2 / m^2$ where $s = n + a$ (31).	1
KierA3	Third alpha modified shape index: $(n\text{-}1)\,(n\text{-}3)^2 / p_3^2$ for odd n, and $(n\text{-}3)\,(n\text{-}2)^2 / p_3^2$ for even n where $s = n + a$ (31).	1
KierFlex	Kier molecular flexibility index: (KierA1) (KierA2) / n (31).	
logs	Log of the aqueous solubility This property is calculated from an atom contribution linear atom type model(34) with $r^2 = 0.90$, ~1,200 molecules.	7
apol	Sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	1
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	
mr	Molecular refractivity (including implicit hydrogens). This property is calculated from an 11 descriptor linear model with $r^2 = 0.997$, RMSE = 0.168 on 1,947 small molecules.	
dipole	Dipole moment calculated from the partial charges of the molecule.	
dipoleX	The x component of the dipole moment (external coordinates).	1
dipoleY	The y component of the dipole moment (external coordinates).	2
dipoleZ	The z component of the dipole moment (external coordinates).	
pmi	Principal moment of inertia.	5
pmiX	x component of the principal moment of inertia (external coordinates).	5
pmiY	y component of the principal moment of inertia (external coordinates).	3
pmiZ	z component of the principal moment of inertia (external coordinates).	3
rgyr	Radius of gyration.	
vsa_acc	Approximation to the sum of VDW surface areas of pure hydrogen bond acceptors (not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH).	
vsa_hyd	Approximation to the sum of VDW surface areas of hydrophobic atoms.	1

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
vsa_pol	Approximation to the sum of VDW surface areas of polar atoms (atoms that are both hydrogen bond donors and acceptors), such as -OH.	
SlogP	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model(29) that calculates logP from the given structure; i.e., the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.	11
SlogP_VSA0	Sum of v_i such that L_i <= -0.4.	
SlogP_VSA2	Sum of v_i such that L_i is in (-0.2,0].	
SlogP_VSA7	Sum of v_i such that L_i is in (0.25,0.30].	
SlogP_VSA8	Sum of v_i such that L_i is in (0.30,0.40].	
SlogP_VSA9	Sum of v_i such that L_i > 0.40.	
SMR	Molecular refractivity (including implicit hydrogens). This property is an atomic contribution model(29) that assumes the correct protonation state (washed structures). The model was trained on ~7000 structures and results may vary from the mr descriptor.	
SMR_VSA2	Sum of v_i such that R_i is in (0.26,0.35].	
SMR_VSA3	Sum of v_i such that R_i is in (0.35,0.39].	
SMR_VSA5	Sum of v_i such that R_i is in (0.44,0.485].	
SMR_VSA7	Sum of v_i such that R_i > 0.56.	
ASA	Water accessible surface area calculated using a radius of 1.4 A for the water molecule. A polyhedral representation is used for each atom in calculating the surface area.	1
ASA+	Water accessible surface area of all atoms with positive partial charge (strictly greater than 0).	
ASA-	Water accessible surface area of all atoms with negative partial charge (strictly less than 0).	
ASA_H	Water accessible surface area of all hydrophobic ($\|q_i\|$<0.2) atoms.	4
ASA_P	Water accessible surface area of all polar ($\|q_i\|$>=0.2) atoms.	1
CASA+	Positive charge weighted surface area, ASA+ times max { q_i > 0 }(35).	
CASA-	Negative charge weighted surface area, ASA- times max { q_i < 0 }(35).	
DASA	Absolute value of the difference between ASA+ and ASA-.	1
DCASA	Absolute value of the difference between CASA+ and CASA-(35).	2

Table 2.4. continued

Code	Descripton	Number of Occurences[b]
FASA+	Fractional ASA+ calculated as ASA+ / ASA.	
FASA-	Fractional ASA- calculated as ASA- / ASA.	
FASA_H	Fractional ASA_H calculated as ASA_H / ASA.	4
FASA_P	Fractional ASA_P calculated as ASA_P / ASA.	4
FCASA+	Fractional CASA+ calculated as CASA+ / ASA.	
FCASA-	Fractional CASA- calculated as CASA- / ASA.	2
VSA	van der Waals surface area. A polyhedral representation is used for each atom in calculating the surface area.	2
TPSA	Polar surface area calculated using group contributions to approximate the polar surface area from connection table information only. The parameterization is that of Ertl *et al.* (36).	
density	Molecular mass density: Weight divided by vdw_vol.	
vdw_area	Area of van der Waals surface calculated using a connection table approximation.	
vdw_vol	van der Waals volume calculated using a connection table approximation.	1
dens	Mass density: molecular weight divided by van der Waals volume as calculated in the vol descriptor.	
glob	Globularity, or inverse condition number (smallest eigenvalue divided by the largest eigenvalue) of the covariance matrix of atomic coordinates. A value of 1 indicates a perfect sphere while a value of 0 indicates a two- or one-dimensional object.	

Table 2.4. continued

std_dim1	Standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	
std_dim2	Standard dimension 2: the square root of the second largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	1
std_dim3	Standard dimension 3: the square root of the third largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	
vol	van der Waals volume calculated using a grid approximation (spacing 0.75 A).	
logP(o/w)	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is calculated from a linear atom type model [LOGP 1998] with $r^2 = 0.931$, RMSE=0.393 on 1,827 molecules.	4

[a] Descriptors searched and their descriptions are generally taken from MOE(7)
[b] Times that the descriptor occurred in the top ranked descriptor combinations

Table 2.5. Top ten "enzyme (FPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *D. discoideum* cell pIC_{50} predictions.

Dictyostelium discoideum (FPPS) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-20.76307 +0.48341 * pIC_{50}Enzyme -14.18055 * GCUT_PEOE_0 +3.20333 * BCUT_PEOE_3	0.69692	1.000000	0.484251	0.729151
-12.32173 +0.57298 * pIC_{50}Enzyme +4.56066 * BCUT_SMR_3 +0.01669 * E_vdw	0.68450	1.000000	0.416094	0.388190
-6.15020 +0.50365 * pIC_{50}Enzyme +3.02716 * BCUT_PEOE_3 +0.000724775 * CASA+	0.67415	1.000000	0.661355	0.393880
-6.94591 +0.49467 * pIC_{50}Enzyme +3.32192 * BCUT_PEOE_3 +0.00390 * ASA+	0.67085	1.000000	0.738924	0.417312
-3.89017 +0.52832 * pIC_{50}Enzyme +2.05150 * BCUT_PEOE_3 +0.01401 * E_vdw	0.66951	1.000000	0.427270	0.353525
2.94011 +0.46041 * pIC_{50}Enzyme -0.00635 * ASA_P +0.63821 * PEOE_PC+	0.66825	1.000000	0.392190	0.638878
-8.79712 +0.50884 * pIC_{50}Enzyme +2.97502 * BCUT_PEOE_3 -0.62648 * PEOE_PC-	0.66574	1.000000	0.643330	0.362862
-17.05598 +0.56036 * pIC_{50}Enzyme +6.27128 * BCUT_SMR_3 +0.000757926 * CASA+	0.66511	1.000000	0.585055	0.370214
-8.75301 +0.48545 * pIC_{50}Enzyme +3.48930 * BCUT_SMR_3 +0.45159 * PEOE_PC+	0.66286	1.000000	0.375754	0.428755
-6.09568 +0.49120 * pIC_{50}Enzyme +3.03006 * BCUT_PEOE_3 +0.35326 * FCASA+	0.66042	1.000000	0.678777	0.361338

Table 2.6. Top ten "enzyme (FPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *L. donovani* cell pIC_{50} predictions.

Leishmania donovani (FPPS) pIC (cell) =	R^{2} [a]	Relative Importance of pIC_{50}(enzyme) [a]	Relative Importance of Descriptor B [a]	Relative Importance of Descriptor C [a]
12.82134 +0.74773 * pIC_{50}Enzyme -1.23653 * FCASA- -14.34579 * Q_VSA_FPOS	0.86952	0.269777	0.915872	1.000000
-1.52445 +0.74773 * pIC_{50}Enzyme -1.23653 * FCASA- +14.34579 * Q_VSA_FNEG	0.86952	0.269777	0.915872	1.000000
2.63995 +0.56184 * pIC_{50}Enzyme +0.10038 * SlogP -2.72841 * b_rotR	0.80829	1.000000	0.615015	0.987653
2.63995 +0.10038 * SlogP -2.72841 * b_1rotR +0.56184 * pIC_{50}Enzyme	0.80829	0.615015	0.987653	1.000000
2.43168 +0.67597 * pIC_{50}Enzyme +0.01062 * PEOE_VSA+0 -4.40941 * Q_VSA_FPOS	0.80745	1.000000	0.806196	0.793464
-1.97773 +0.67597 * pIC_{50}Enzyme +0.01062 * PEOE_VSA+0 +4.40940 * Q_VSA_FNEG	0.80745	0.793464	0.806196	1.000000
2.78569 +0.67570 * pIC_{50}Enzyme -3.01289 * Q_VSA_FPOS +0.14382 * SlogP	0.80578	1.000000	0.861479	0.732693
-0.22720 +0.67570 * pIC_{50}Enzyme +0.14382 * SlogP +3.01289 * Q_VSA_FNEG	0.80578	1.000000	0.732693	0.861479
5.23713 +0.35909 * pIC_{50}Enzyme +0.00796 * Q_VSA_HYD -0.19768 * KierA1	0.80375	0.319957	0.726844	1.000000
4.83355 +0.30356 * pIC_{50}Enzyme -0.22903 * KierA1 +0.00805 * VSA	0.80300	0.233458	1.000000	0.738323

[a] Identical values are obtained when descriptors are linearly dependent

Table 2.7. Top ten "enzyme (UPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *S. pneumoniae* cell pIC_{50} predictions.

Streptococcus pneumoniae (UPPS) pIC (cell) =	R^2	Relative Importance of pIC_{50}Enzyme	Relative Importance of Descriptor B	Relative Importance of Descriptor C
2.36039 +0.09521 * pIC_{50}Enzyme +5.57167 * E_oop +0.30641 * logS	0.70859	0.206740	0.770366	1.000000
2.73916 -0.02058 * pIC_{50}Enzyme -0.25615 * dipoleY -0.50252 * SlogP	0.65294	0.036773	0.664638	1.000000
4.02259 -0.10469 * pIC_{50}Enzyme +0.44931 * logS -0.00111 * pmiZ	0.63628	0.155024	1.000000	0.380149
1.93265 +0.02581 * pIC_{50}Enzyme +5.82244 * E_oop -0.38762 * SlogP	0.63029	0.059775	0.858685	1.000000
2.91917 +0.05651 * pIC_{50}Enzyme -0.16171 * dipoleY +0.33943 * logS	0.61395	0.110774	0.460388	1.000000
3.29294 +0.11608 * pIC_{50}Enzyme +8.55584 * E_oop -0.01199 * vsa_hyd	0.60705	0.213066	1.000000	0.682444
3.00970 -0.02670 * pIC_{50}Enzyme +0.30685 * logS -0.000145748 * pmiY	0.59586	0.057891	1.000000	0.472137
2.72405 +0.01829 * pIC_{50}Enzyme +0.01022 * E_nb +0.37898 * logS	0.59540	0.032109	0.334330	1.000000
0.97384 +0.21900 * pIC_{50}Enzyme -0.000223701 * pmiY +6.34230 * E_oop	0.59461	0.542273	0.827566	1.000000
2.46641 +0.03924 * pIC_{50}Enzyme +0.12363 * E_str +0.40087 * logS	0.59423	0.065125	0.332503	1.000000

Table 2.8. Top ten "enzyme (MurI) plus 2-descriptor" combinations with their coefficients and relative contributions for the *S. pneumoniae* cell pIC_{50} predictions.

Streptococcus pneumoniae (MurI) pIC (cell) =	R^{2} [a]	Relative Importance of pIC_{50}(enzyme) [a]	Relative Importance of Descriptor B [a]	Relative Importance of Descriptor C [a]
10.52552 +0.67868 * pIC_{50}Enzyme -8.00526 * FASA_H -9.58941 * Q_VSA_FPOL	0.72033	0.760086	0.930312	1.000000
2.52026 +0.67868 * pIC_{50}Enzyme -9.58941 * Q_VSA_FPOL +8.00526 * FASA_P	0.72033	0.760086	1.000000	0.930312
0.93611 +0.67868 * pIC_{50}Enzyme -8.00526 * FASA_H +9.58941 * Q_VSA_FHYD	0.72033	0.760086	0.930312	1.000000
-7.06915 +0.67868 * pIC_{50}Enzyme +9.58941 * Q_VSA_FHYD +8.00526 * FASA_P	0.72033	0.760086	1.000000	0.930312
1.11064 +0.73485 * pIC_{50}Enzyme +0.00116 * pmiZ -0.00189 * DCASA	0.70598	1.000000	0.352718	0.412965
0.96166 +0.74218 * pIC_{50}Enzyme -0.01333 * ASA_H +0.02650 * Q_VSA_HYD	0.70104	0.683156	1.000000	0.978195
1.07105 +0.73955 * pIC_{50}Enzyme +0.00121 * pmiZ -0.00932 * DASA	0.69820	1.000000	0.367214	0.401603
5.68232 +0.68410 * pIC_{50}Enzyme -5.36646 * Q_VSA_FPOL -0.00565709 * ASA_H	0.69080	1.000000	0.730430	0.673985
0.31586 +0.68410 * pIC_{50}Enzyme -0.00565708 * ASA_H +5.36646 * Q_VSA_FHYD	0.69080	1.000000	0.673985	0.730430
-0.26123 +0.77505 * pIC_{50}Enzyme -0.00154943 * DCASA +10.54584 * PEOE_RPC+	0.68937	1.000000	0.261402	0.321590

[a] Identical values are obtained when descriptors are linearly dependent

Table 2.9. Top ten "enzyme (NS3 protease) plus 2-descriptor" combinations with their coefficients and relative contributions for the HCV replicon cell pIC_{50} predictions.

Hepatitis C Virus (NS3 protease) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-1.29239 +0.91301 * pIC_Enzyme +0.58847 * log(PAMPA) -0.00017357 * pmiY	0.77744	0.730339	1.000000	0.305259
-2.44335 +0.92235 * pIC_{50}Enzyme +0.55294 * log(PAMPA) +0.25772 * log(Caco-2)	0.74101	0.785205	1.000000	0.261151
0.11561 +0.84966 * pIC_{50}Enzyme +0.55981 * log(PAMPA) -0.56438 * std_dim2	0.73967	0.714448	1.000000	0.231869
-2.11082 +0.91600 * pIC_{50}Enzyme +0.58198 * log(PAMPA) -0.16375 * dipoleX	0.73819	0.740897	1.000000	0.222535
-1.07166 +0.92801 * pIC_{50}Enzyme +0.62367 * log(PAMPA) -0.24639 * PEOE_PC+	0.73569	0.700433	1.000000	0.198231
-1.43622 +0.92293 * pIC_{50}Enzyme +0.65322 * log(PAMPA) -0.06097 * KierA2	0.73564	0.665088	1.000000	0.197280
-1.52542 +0.92611 * pIC_{50}Enzyme +0.64820 * log(PAMPA) -0.04420 * Kier2	0.73518	0.672545	1.000000	0.195686
-1.58484 +0.91365 * pIC_{50}Enzyme +0.65058 * log(PAMPA) -0.05945 * Kier3	0.73454	0.661071	1.000000	0.193573
-1.85996 +0.96397 * pIC_{50}Enzyme +0.61988 * log(PAMPA) -0.00374 * Q_VSA_PPOS	0.73415	0.732031	1.000000	0.203505
-1.61354 +0.95796 * pIC_{50}Enzyme +0.61521 * log(PAMPA) -0.00310 * Q_VSA_POL	0.73394	0.732983	1.000000	0.203058

Table 2.10. Top ten "enzyme (HIV-1 integrase) plus 2-descriptor" combinations with their coefficients and relative contributions for the HIV-1 integrase replicon cell pIC_{50} predictions.

MT4 cells (HIV-1 integrase) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-4.83173 +0.98896 * pIC50_enzyme +1.36334 * pCC50 -0.00697 * Weight	0.68808	0.892564	1.000000	0.526629
-3.90054 +0.79998 * pIC50_enzyme +1.24225 * pCC50 -0.00990 * vsa_hyd	0.68552	0.792381	1.000000	0.520038
-5.92926 +0.89119 * pIC50_enzyme +1.37520 * pCC50 -0.000147 * pmi	0.68063	0.797385	1.000000	0.489095
-5.67093 +0.84883 * pIC50_enzyme +1.35713 * pCC50 -0.000166 * pmiX	0.67775	0.769595	1.000000	0.482305
-2.59105 +0.71970 * pIC50_enzyme +1.22060 * pCC50 -0.30594 * chi1v	0.67651	0.725509	1.000000	0.509833
-5.53671 +0.85017 * pIC50_enzyme +1.34149 * pCC50 -0.19817 * dipole	0.67241	0.779798	1.000000	0.477547
-5.45844 +0.84242 * pIC50_enzyme +1.33341 * pCC50 -0.19829 * dipoleX	0.67078	0.777374	1.000000	0.475734
-2.07656 +0.67785 * pIC50_enzyme +1.16006 * pCC50 -0.00792 * vol	0.66766	0.718981	1.000000	0.519278
-2.88642 +0.78860 * pIC50_enzyme +1.21043 * pCC50 -0.18150 * chi0v	0.66619	0.801639	1.000000	0.497274
-6.15351 +0.91486 * pIC50_enzyme +1.35583 * pCC50 +0.20861 * logS	0.66278	0.830251	1.000000	0.471726

Table 2.11. Top ten "enzyme (hlGP) plus 2-descriptor" combinations with their coefficients and relative contributions for the rat hepatocytes pIC_{50} predictions.

Rat Hepatocytes (hlGP) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-2.48325 +0.74485 * pIC_{50}Enzyme +0.02112 * PEOE_VSA_PNEG +0.00797 * Q_VSA_HYD	0.71312	1.000000	0.671022	0.456010
-2.61987 +0.81722 * pIC_{50}Enzyme +0.01567 * PEOE_VSA_PNEG +0.38040 * E_str	0.70675	1.000000	0.453849	0.360041
-2.47800 +0.77133 * pIC_{50}Enzyme +0.02157 * PEOE_VSA_PNEG +0.00356 * ASA_H	0.70330	1.000000	0.661853	0.414954
-3.30393 +0.77214 * pIC_{50}Enzyme +0.02442 * PEOE_VSA_PNEG +3.06864 * FASA_H	0.70198	1.000000	0.748269	0.455019
-0.23529 +0.77214 * pIC_{50}Enzyme +0.02442 * PEOE_VSA_PNEG -3.06864 * FASA_P	0.70198	1.000000	0.748269	0.455019
-7.57118 +0.89039 * pIC_{50}Enzyme +0.01755 * PEOE_VSA_PNEG -2.55393 * BCUT_PEOE_0	0.70019	1.000000	0.466411	0.329457
-8.83556 +0.83245 * pIC_{50}Enzyme +0.01546 * PEOE_VSA_PNEG +2.90552 * GCUT_SMR_3	0.69768	1.000000	0.439538	0.327327
-7.31859 +2.61295 * GCUT_SLOGP_3 +0.79188 * pIC_{50}Enzyme +0.01647 * PEOE_VSA_PNEG	0.69753	1.000000	0.345396	0.492133
-7.31859 +0.79188 * pIC_{50}Enzyme +0.01647 * PEOE_VSA_PNEG +2.61295 * GCUT_SLOGP_3	0.69753	1.000000	0.492133	0.345396
-2.00842 +0.73115 * pIC_{50}Enzyme +0.01897 * PEOE_VSA_PNEG +0.00634 * vsa_hyd	0.69664	1.000000	0.614097	0.397957

Table 2.12. Top ten "enzyme (KDR kinase) plus 2-descriptor" combinations with their coefficients and relative contributions for the NIH3T3 cell pIC_{50} predictions.

NIH 3T3 (KDR, type III receptor tyrosine kinase) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
0.82478 +0.85331 * pIC_{50}Enzyme +6.56543 * PEOE_VSA_FPOL +0.05504 * E_tor	0.60006	1.000000	0.494191	0.718790
0.66511 +0.86124 * pIC_{50}Enzyme +0.05066 * E_tor +12.60955 * PEOE_VSA_FPNEG	0.59907	1.000000	0.655445	0.485462
9.32007 +0.79676 * pIC_{50}Enzyme -18.85532 * PEOE_RPC- -0.06827 * apol	0.59487	0.766096	0.760710	1.000000
10.76167 +0.82694 * pIC_{50}Enzyme -19.03382 * PEOE_RPC- -0.00815127 * ASA	0.59352	0.795265	0.768062	1.000000
10.23883 +0.80199 * pIC_{50}Enzyme -20.80145 * PEOE_RPC- -0.00895460 * vdw_vol	0.59328	0.730389	0.794900	1.000000
10.31493 +0.79311 * pIC_{50}Enzyme -20.41614 * PEOE_RPC- -0.01264080 * vol	0.59276	0.729847	0.788320	1.000000
10.31493 +0.79311 * pIC_{50}Enzyme -20.41614 * PEOE_RPC- -0.01264080 * vol	0.58789	0.708545	0.798883	1.000000
-0.21236 +0.91159 * pIC_{50}Enzyme +0.03452 * E -0.05235 * E_nb	0.58380	0.500955	0.743845	1.000000
1.32645 +0.89890 * pIC_{50}Enzyme -0.01124 * E_nb +0.04199 * E_tor	0.58330	1.000000	0.434552	0.520573
2.55668 +0.72281 * pIC_{50}Enzyme -0.75267 * KierA3 +1.19503 * PEOE_PC+	0.58114	0.783185	1.000000	0.775189

Table 2.13. Top ten "enzyme (Akt kinase) plus 2-descriptor" combinations with their coefficients and relative contributions for the MiaPaCa-2 cell pIC_{50} predictions.

MiaPaca-2 human pancreatic cancer cells (Akt Kinase) pIC (cell) =	R^2	Relative Importance of pIC_{50}Enzyme	Relative Importance of Descriptor B	Relative Importance of Descriptor C
4.33712 +0.45265 * pIC_{50}Enzyme -0.0000929650 * pmiX -4.23429 * GCUT_SLOGP_2	0.68326	1.000000	0.556839	0.424186
2.55169 +0.44882 * pIC_Enzyme -0.000099391 * pmi +0.00253 * ASA_H	0.68020	1.000000	0.711354	0.499224
4.44515 +0.39949 * pIC_{50}Enzyme -0.0000873893 * pmiX -0.02504 * E_ele	0.67971	1.000000	0.593093	0.458662
4.23090 +0.42619 * pIC_{50}Enzyme -0.0000879041 * pmiX -1.87949 * FASA_P	0.67557	1.000000	0.559220	0.423219
2.35141 +0.42619 * pIC_{50}Enzyme -0.0000879041 * pmiX +1.87949 * FASA_H	0.67557	1.000000	0.559220	0.423219
4.65687 +0.41086 * pIC_{50}Enzyme -0.0000778479 * pmi -4.34328 * GCUT_SLOGP_2	0.67513	1.000000	0.608637	0.479355
4.54500 +0.38640 * pIC_{50}Enzyme -1.96877 * FASA_P -0.0000739953 * pmi	0.67115	1.000000	0.488967	0.615143
2.57623 +0.38640 * pIC_{50}Enzyme -0.0000739953 * pmi +1.96877 * FASA_H	0.67115	1.000000	0.615143	0.488967
4.34247 +0.40352 * pIC_Enzyme -0.0025494 * ASA_P -0.0000831233 * pmiX	0.67007	1.000000	0.435308	0.558505
4.71925 +0.35970 * pIC_{50}Enzyme -0.0000719880 * pmi -0.02523 * E_ele	0.66794	1.000000	0.642875	0.513363

Figure 2.1. Structures of the 26 compounds investigated in cell (*Plasmodium falciparum*) and enzyme (*Plasmodium vivax* FPPS) assays.

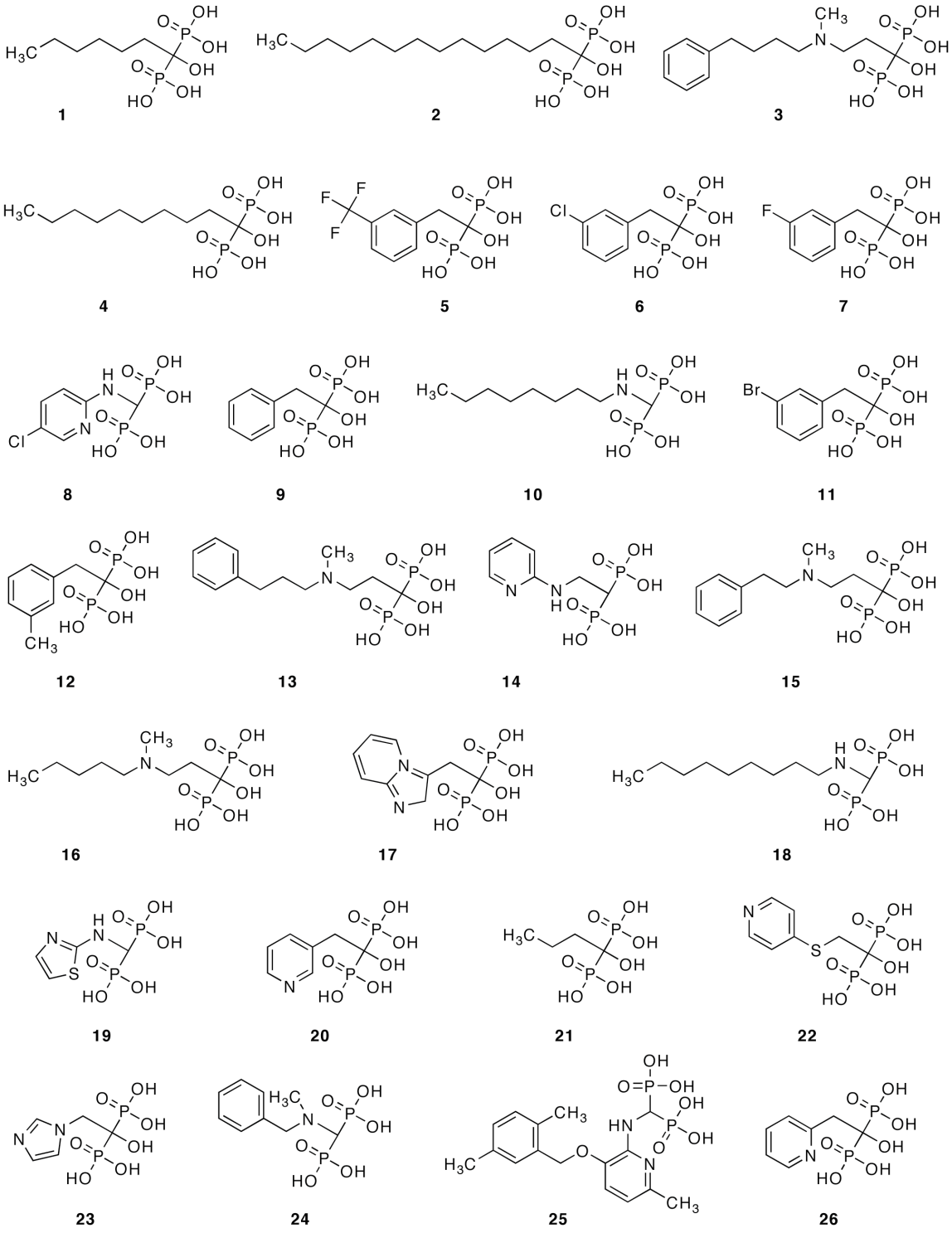


Figure 2.2. Correlation plots for cell (*Plasmodium falciparum*) and enzyme (*Plasmodium vivax* FPPS) assays and predicted cell activities from the training and test set data, obtained by using the combinatorial descriptor search method. (a) Plot showing correlation between cell pIC_{50} (= - $\log_{10}\{IC_{50}\}$) and enzyme pIC_{50} values. (b) Best correlation between predicted cell pIC_{50} (enzyme plus two molecular descriptors) and experimental pIC_{50} values: training set results. (c) Test set pIC_{50} predictions (leave-two-out analysis) plotted against the experimental values. The R^2 value increases from ~0 to 0.74, when adding the two molecular descriptors to the enzyme data. The colored circles (A-C) indicate bisphosphonates with side chains having different relative hydrophobicities and potencies in the cell assay (low, intermediate or high, respectively), as discussed in the Text.

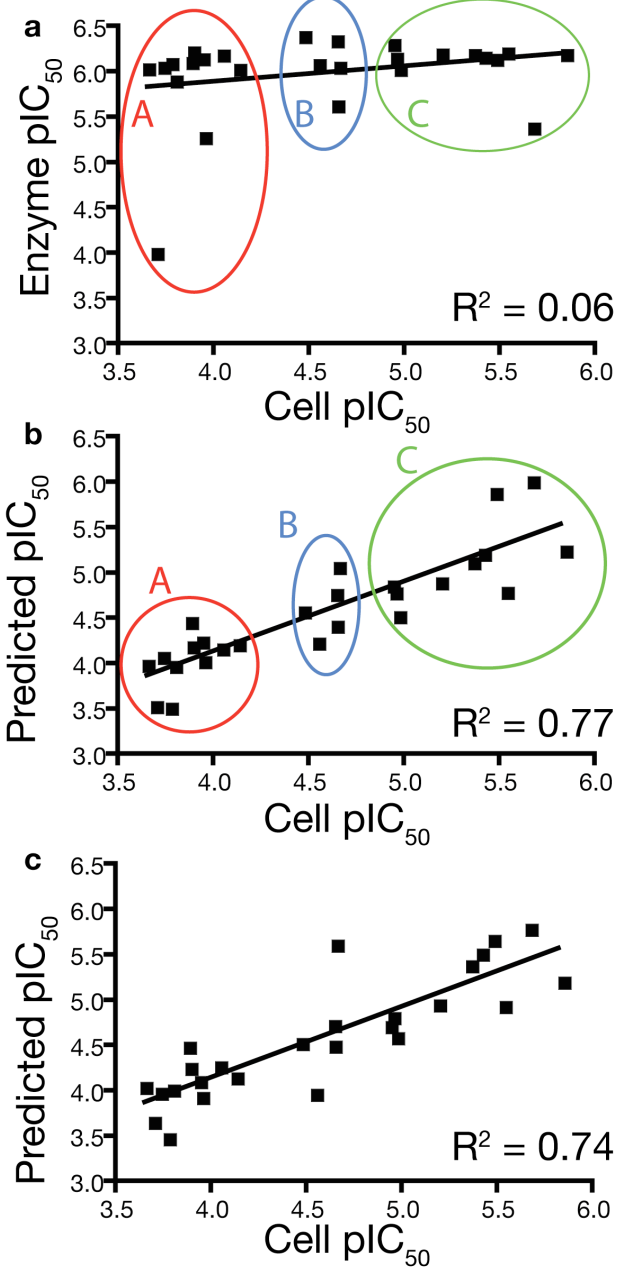


Figure 2.3. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in: *Dictyostelium discoideum* (a-c), *Leishmania donovani* (d-f) and *Streptococcus pneumoniae* (g-i). (a) Plot showing the correlation between cell (*D. discoideum*) pIC_{50} and enzyme (FPPS) pIC_{50} values. (b) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (c) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (d) Plot showing the correlation between cell (*L. donovani*) pIC_{50} and enzyme (FPPS) pIC_{50} values. (e) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (f) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (g) Plot showing the correlation between cell (*S. pneumoniae*) pIC_{50} and enzyme (UPPS) pIC_{50} values. (h) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (i) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis.

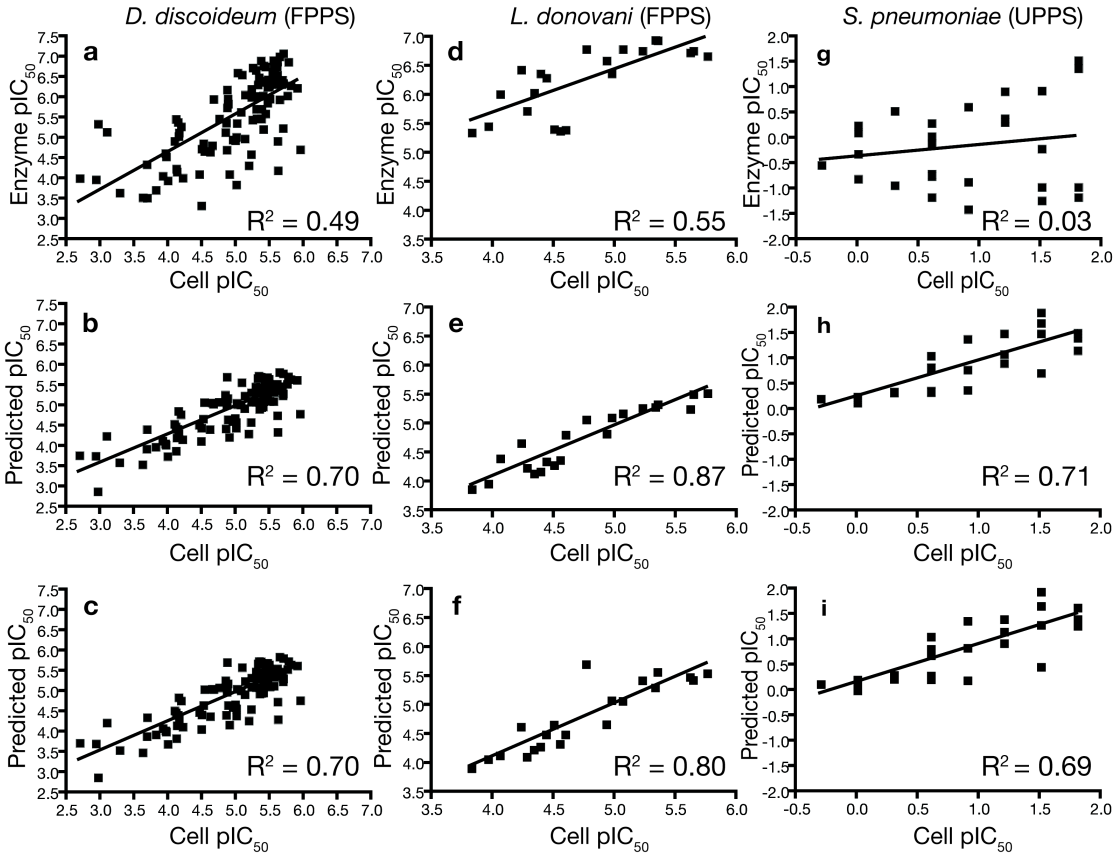


Figure 2.4. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in *Streptococcus pneumoniae* (MurI) (a-c), Hu-7 (HCV NS3 protease replication) (d-f) and MT4 (HIV-1 integrase) (g-i). (a) Plot showing the correlation between cell (*S. pneumoniae*) pIC_{50} and enzyme (MurI) pIC_{50} values. (b) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (c) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (d) Plot showing the correlation between cell (HCV/Hu-7) pIC_{50} and enzyme (HCV NS3 protease) pIC_{50} values. (e) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (f) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (g) Plot showing the correlation between cell (HIV/MT4) pIC_{50} and enzyme (HIV-1 integrase) pIC_{50} values. (h) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (i) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis.

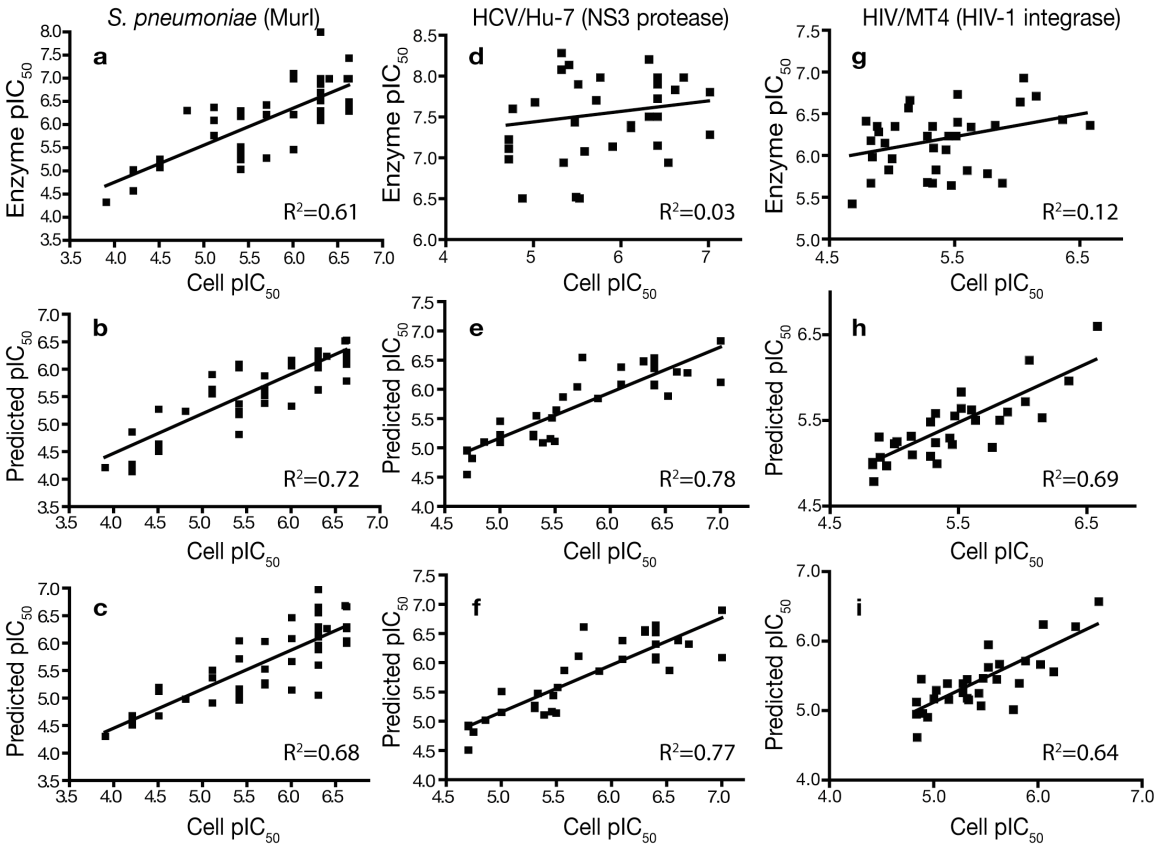
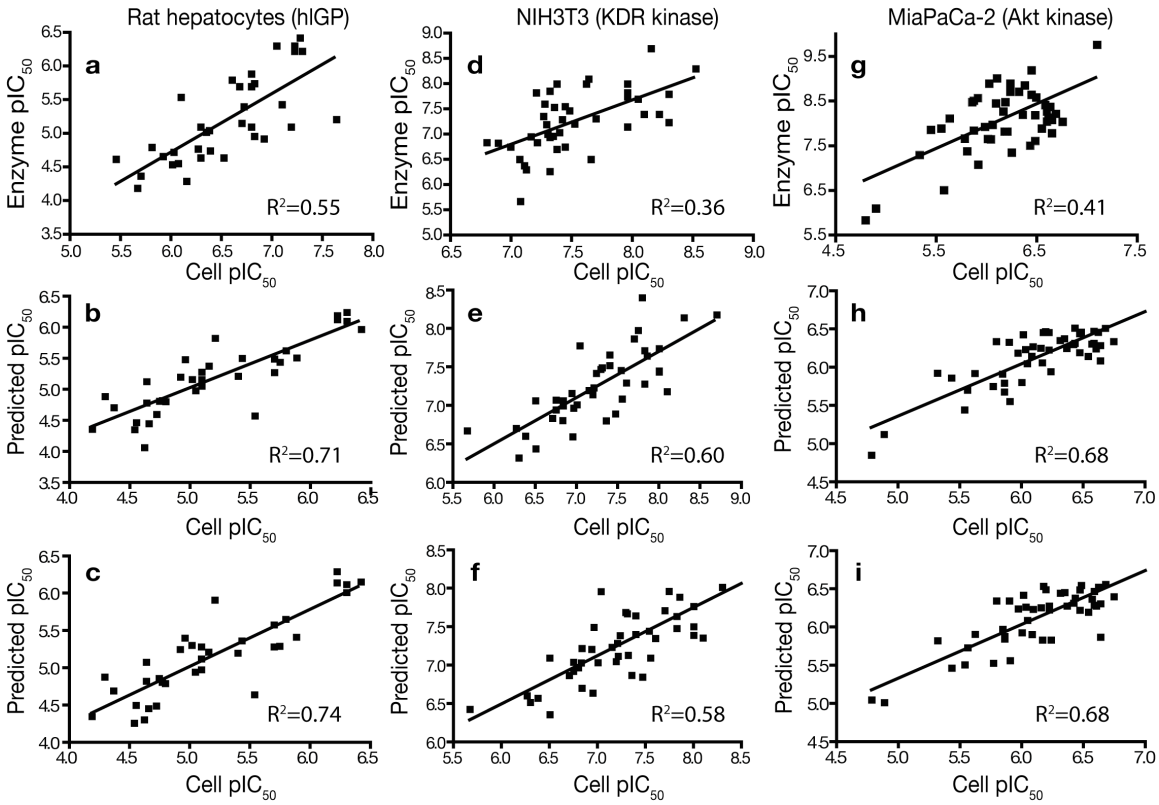


Figure 2.5. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in rat hepatocytes (hlGP) (a-c), NIH3T3 (KDR kinase) (d-f) and MiaPaCa-2 (Akt kinase) (g-i). (a) Plot showing the correlation between cell (Rat hepatocytes) pIC_{50} and enzyme (hlGP) pIC_{50} values. (b) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (c) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (d) Plot showing the correlation between cell (NIH 3T3) pIC_{50} and enzyme (KDR kinase) pIC_{50} values. (e) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (f) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (g) Plot showing the correlation between cell (MiaPaCa-2) pIC_{50} and enzyme (Akt kinase) pIC_{50} values. (h) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (i) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis.



2.7 References

1. Buckner FS, Eastman RT, Yokoyama K, Gelb MH, & Van Voorhis WC (2005) Protein Farnesyl Transferase Inhibitors for the Treatment of Malaria and *African trypanosomiasis*. *Curr Opin Investig Drugs* 6(8):791-797.
2. Dunford JE*, et al.* (2008) Structure-activity relationships among the nitrogen containing bisphosphonates in clinical use and other analogues: time-dependent inhibition of human farnesyl pyrophosphate synthase. *J Med Chem* 51(7):2187-2195.
3. Martin MB*, et al.* (2001) Bisphosphonates inhibit the growth of Trypanosoma brucei, Trypanosoma cruzi, Leishmania donovani, Toxoplasma gondii, and Plasmodium falciparum: a potential route to chemotherapy. *J Med Chem* 44(6):909-916.
4. Ghosh S*, et al.* (2004) Effects of Bisphosphonates on the Growth of *Entamoeba histolytica* and *Plasmodium* Species *In vitro* and *In vivo*. (Translated from English) *J. Med. Chem.* 47(1):175-187 (in English).
5. Dobson PD & Kell DB (2008) Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat Rev Drug Discov* 7(3):205-220.
6. Peukert S*, et al.* (2008) Design and structure-activity relationships of potent and selective inhibitors of undecaprenyl pyrophosphate synthase (UPPS): tetramic, tetronic acids and dihydropyridin-2-ones. *Bioorg Med Chem Lett* 18(6):1840-1844.
7. Anonymous (2006) Molecular Operating Environment (MOE) (Chemical Computing Group, Inc., Montreal, Quebec).

8. Szabo CM, Martin MB, & Oldfield E (2002) An Investigation of Bone Resorption and *Dictyostelium discoideum* Growth Inhibition by Bisphosphonate Drugs. *J Med Chem* 45(14):2894-2903.

9. Yardley V, *et al.* (2002) In vivo activities of farnesyl pyrophosphate synthase inhibitors against Leishmania donovani and Toxoplasma gondii. *Antimicrob Agents Chemother* 46(3):929-931.

10. Montalvetti A, *et al.* (2003) Farnesyl pyrophosphate synthase is an essential enzyme in Trypanosoma brucei. In vitro RNA interference and in vivo inhibition studies. *J Biol Chem* 278(19):17075-17083.

11. Montalvetti A, *et al.* (2001) Bisphosphonates are Potent Inhibitors of *Trypanosoma cruzi* Farnesyl Pyrophosphate Synthase. *J Biol Chem* 276(36):33930-33937.

12. Sanders JM, *et al.* (2003) 3-D QSAR Investigations of the Inhibition of *Leishmania major* Farnesyl Pyrophosphate Synthase by Bisphosphonates. *J Med Chem* 46(24):5171-5183.

13. Dunford JE, *et al.* (2001) Structure-Activity Relationships for Inhibition of Farnesyl Diphosphate Synthase *In vitro* and Inhibition of Bone Resorption *In vivo* by Nitrogen-Containing Bisphosphonates. *J Pharmacol Exp Ther* 296(2):235-242.

14. Fujihashi M, *et al.* (2001) Crystal Structure of *cis*-Prenyl Chain Elongating Enzyme, Undecaprenyl Diphosphate Synthase. *Proc Natl Acad Sci U S A* 98(8):4337-4342.

15. Guo RT, *et al.* (2007) Bisphosphonates target multiple sites in both cis- and trans-prenyltransferases. *Proc Natl Acad Sci U S A* 104(24):10022-10027.

16. de Dios A, *et al.* (2002) 4-Substituted D-glutamic acid analogues: the first potent inhibitors of glutamate racemase (MurI) enzyme with antibacterial activity. *J Med Chem* 45(20):4559-4570.

17. Li C, *et al.* (2008) Correlation between PAMPA permeability and cellular activities of hepatitis C virus protease inhibitors. *Biochem Pharmacol* 75(5):1186-1197.

18. Verschueren WG, *et al.* (2005) Design and optimization of tricyclic phtalimide analogues as novel inhibitors of HIV-1 integrase. *J Med Chem* 48(6):1930-1940.

19. Klabunde T, *et al.* (2005) Acyl ureas as human liver glycogen phosphorylase inhibitors for the treatment of type 2 diabetes. *J Med Chem* 48(20):6178-6193.

20. Dinges J, *et al.* (2007) 1,4-Dihydroindeno[1,2-c]pyrazoles with acetylenic side chains as novel and potent multitargeted receptor tyrosine kinase inhibitors with low affinity for the hERG ion channel. *J Med Chem* 50(9):2011-2029.

21. Zhu GD, *et al.* (2007) Syntheses of potent, selective, and orally bioavailable indazole-pyridine series of protein kinase b/akt inhibitors with reduced hypotension. *J Med Chem* 50(13):2990-3003.

22. Anonymous (2005) *GraphPad Prism Version 4.0c* (GraphPad Software, Inc., La Jolla, CA).

23. Anonymous (2007) *MATLAB Version 7.4* (The Mathworks, Inc., Natick, MA).

24. Anonymous (Sybyl 7.3 (Tripos, Inc., St. Louis, MO).

25. Petitjean M (1992) Applications of the Radius Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical-Compounds. (Translated from English) *Journal of Chemical Information and Computer Sciences* 32(4):331-337 (in English).

26. Balaban AT (1979) Five New Topological Indices for the Branching of Tree-Like Graphs. *Theoretica Chimica Acta.* 53:355-375.

27. Weiner H (1947) Structural Determination of Paraffin Boiling Points. *J Am Chem Soc* 69:17-20.

28. Pearlman RS & Smith KM (1998) Novel Software Tools for Chemical Diversity. *Persp Drug Disc Desc 9/10/11*:339-353.

29. Wildman SA & Crippen GM (1999) Prediction of physicochemical parameters by atomic contributions. (Translated from English) *Journal of Chemical Information and Computer Sciences* 39(5):868-873 (in English).

30. Hall LH & Kier LB (1977) The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity. *Eur J Med Chem* 12:307.

31. Hall LH & Kier LB (1991) The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling. *Reviews of Computational Chemistry* 2.

32. Balaban AT (1982) Highly Discriminating Distance-Based Topological Index. *Chem Phys Lett* 89(5):399-404.

33. Oprea TI (2000) Property Distribution of Drug-Related Chemical Databases. *J Comp Aid Mol Des* 14:251-264.

34. Hou TJ, Xia K, Zhang W, & Xu XJ (2004) ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J Chem Inf Comput Sci* 44:266-275.

35. Stanton D & Jurs P (1990) *Anal Chem* 62:2323.

36. Ertl P, Rohde B, & Selzer P (2000) Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* 43:3714-3717.

Chapter 3

Lipophilic Analogs of Zoledronate and Risedronate Inhibit *Plasmodium* Geranylgeranyl Diphosphate Synthase and Have Potent Activity Against Malaria Parasites

3.1 Notes and Acknowledgements

This work was supported by the United States Public Health Science (NIH grant GM 65307). High throughput screening (HTS) assay, in vitro cell assay and in vivo experiment were carried out by Dr. Fernando Dossin at Institut Pasteur Korea.

3.2 Introduction

Malaria, caused by *Plasmodium* spp., causes ~1 million deaths each year (1) and there are ever present problems due to drug resistance (2). There is, therefore, a need for new drugs, and drug leads. In earlier work, we and others found that the bisphosphonate class of drugs (3) used to treat bone-related diseases-osteoporosis, Paget's disease and hypercalcemia due to malignancy, also inhibited the growth of a range of parasitic protozoa including *Trypanosoma cruzi*(4, 5), *T. brucei* (4, 6), *Leishmania spp*. (4, 7, 8), *Toxoplasma gondii* (4, 9), *Cryptosporidium parvum* (10, 11), *Entamoeba histolytica* (4, 12, 13) and *Plasmodium* spp. (4, 13-15). In the case of *Plasmodium* spp., the most potent inhibitors were not, however, the nitrogen-containing bisphosphonates such as risedronate or zoledronate used to treat bone diseases, but more lipophilic n-alkyl bisphosphonates (13), although their target in *P. falciparum* was not

determined. More recently, a *P. vivax* geranylgeranyl diphosphate synthase (PvGGPPS) has been cloned, expressed, purified and crystallized and its three-dimensional structure determined (16). The enzyme is potently inhibited by bisphosphonates (16) so it seemed possible that it might be a target for the inhibitors discovered earlier. To investigate this possibility, we recently determined the IC_{50} values for 25 bisphosphonates against PvGGPPS and compared the results for enzyme inhibition with *P. falciparum* growth inhibition pIC_{50} (= $-\log_{10}IC_{50}$) values (17). The correlation was very poor: $R^2 = 0.06$. For example, zoledronate had an IC_{50} of ~790 nM in the enzyme assay, but an IC_{50} = 120 μM in cells, and similar results were found with several other systems with the average R^2 value being 0.30 (17) for 10 diverse assays (3 from our group; 7 from other groups). This strongly suggested that zoledronate had poor permeability in the *P. falciparum*/red cell assay, due to its highly polar nature (log P = -2.9). However, when incorporating mathematical descriptors (such as logP) to begin to account for permeability, we found that good (R^2=0.7) correlations between cell and enzyme activity could be obtained (17). These results indicated, at least for these types of compounds, that simply screening for good enzyme inhibitors might not be particularly informative since many such inhibitors would be unable to get into cells, and that cell assays would be much more desirable.

A second set of problems with the bisphosphonate class of molecules is that they bind very tightly to bone mineral (18, 19), resulting in their rapid removal from the bloodstream. This is of course a desirable feature of a bone drug, but not of an anti-infective (or anti-cancer) drug lead and in recent work, we have been developing a class of compounds called "lipophilic bisphosphonates" (20, 21) in which the 1-OH group on the bisphosphonate backbone, part of the tri-dentate "bone-hook", is removed and in which a variety of hydrophobic side-chains are attached to the molecules in order to increase clogP values, typically from ~ -2 or -3 to ~ 2 or 3.

These lipophilic bisphosphonates have far more potent activity both *in vitro* and *in vivo* than do conventional bisphosphonates in tumor cell growth inhibition and γδ T cell activation assays (20, 21), as well as against malaria parasites (14). In this work, we thus elected to screen our in-house library of 564 prenyl synthase inhibitors, developed over the past decade as anti-cancer drug leads and as anti-bacterials (20-24), for their activity in *P. falciparum* growth inhibition inside red cells. We discovered two potent leads, lipophilic analogs of the commercial drugs risedronate and zoledronate and determined their single crystal x-ray crystallographic structures bound to *P. vivax* GGPPS, in addition to testing them *in vivo*, finding potent activity, opening up the possibility that other commercial bisphosphonates, inactive themselves against *Plasmodium* spp. as well as other organisms) may be converted to species that are highly active both *in vitro* and *in vitro*, via simple chemical modification.

3.3 Results and Discussion

Inhibitor Library and High Throughput Screening (HTS).

We screened 564 compounds for their ability to inhibit *Plasmodium falciparum* cell growth inside erythrocyctes. The library consisted of a diverse selection of bisphosphonates, phosphonosulfonates and cationic species known to inhibit head-to-tail and/or head-to-head prenyl transferases that have been developed in our laboratory over the past decade, primarily as anti-cancer and anti-bacterial agents.

In *Plasmodium* spp., the initial steps in isoprenoid biosynthesis are carried out by the so-called methylerythritol phosphate (MEP) pathway, which produces isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) from pyruvate and glyceraldehyde-3-phosphate, Figure 3.1 (25). DMAPP then condenses sequentially with three molecules of IPP to form geranyl

diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), which is then used to prenylate proteins (26). In addition, in *P. falciparum*, GGPP is converted via prephytoene diphosphate to phytoene and thence to carotenoids (27), plus, the longer chain diphosphates are converted to quinones such as Men-4 (28) as well as dolichols (29). In *Plasmodium* spp., GGPPS appears to be bi-functional, making both FPP as well as GGPP, but based on its sequence and x-ray crystallographic structure with bound inhibitors (16), the enzyme is clearly more similar to other FPPS than GGPPS enzymes – as evidenced for example by the presence of a third Asp in the second conserved DDXXD-domain (16), inhibition by zoledronate and risedronate (unlike human and *S. cerevisae* GGPPS), and the presence of the $3Mg^{2+}$ seen in these structures, compared to the $2Mg^{2+}$ typically found in *S. cerevisae* GGPPS (30). However, the enzyme is a more effective producer of GGPP than of FPP (16). Our library contained 474 bisphosphonates that might inhibit the *Plasmodium* GGPPS, in addition to 5 phosphonosulfonates and related systems that inhibit head-to-head prenyl synhtases (22-24) such as the *Staphylococcus aureus* protein CrtM (dehydrosqualene; (22)). Such compounds inhibit both dehydrosqualene as well as squalene synthase (31) and could also inhibit *Plasmodium* phytoene synthase. We also screened 34 cationic species (such as Ro 48-8071 and quinuclidines), some of which are known to inhibit *Plasmodium* cell growth (28, 32). In addition, we screened 51 pyrophosphate mimics that are potential inhibitor in isoprenoid pathway.

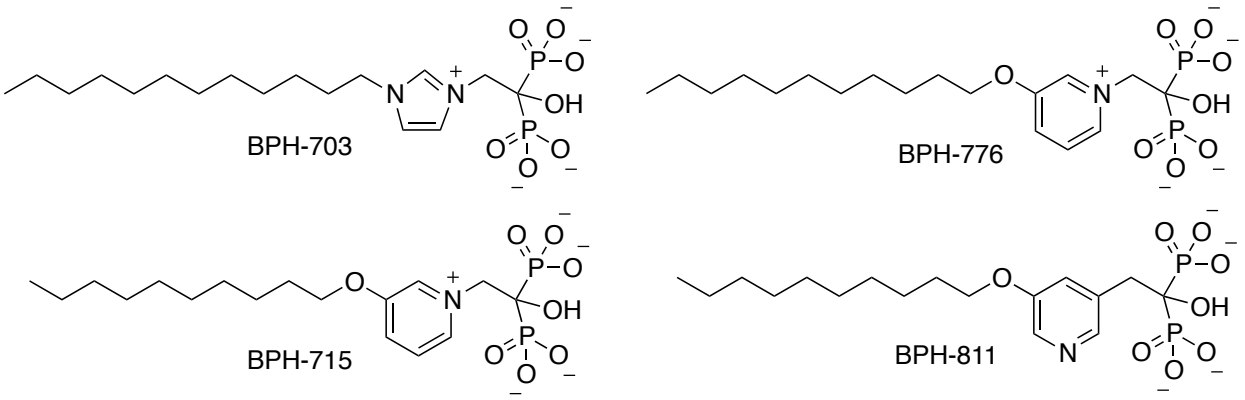
We first screened all compounds at 10 μM to find possible hits, using artemisinin as a positive control (IC_{50} = 20 nM). There were 78 hits (defined as giving ≥ 70% inhibition at 10 μM), Figure 3.2A, and the Z′ factor for the control wells was 0.72, Figure 3.2B. The Z′ factor is defined (33) as:

$$Z' = 1 - 3(\sigma_c^+ + \sigma_c^-) / | \mu_c^+ - \mu_c^- |$$

where σ_c^+/σ_c^- are the standard deviations of the positive (artemisinin)/negative (phosphate buffered saline) controls and μ_c^+/μ_c^- are the corresponding mean values (33) with $Z' = 0.72$ indicating an excellent assay. Full assay details are given in Methods. We then determined dose-response curves for these hits (structures shown in Figure 3.7): dose-response curves are shown in Figure 3.8. The chemical structures of the 10 most active compounds are shown in Figure 3.9. To determine which of these compounds might be worth further investigation, we next determined the IC_{50} values for each of these 10 compounds in inhibiting the growth of three human cell lines. Typical dose-response curves are shown in Figure 3.2C and all IC_{50} values are given in Table 3.2. We then used a therapeutic index plot, Figure 3.2D, in which we plotted the therapeutic index (T.I.), defined as:

T.I. (therapeutic index) = IC_{50} (human cell assay)/IC_{50} (*P. falciparum* assay)

versus the *P. falciparum* IC_{50} values to find those compounds with the best activity versus *P. falciparum* that also had the low toxicity to human cells. Most compounds were very toxic but two: BPH-703 and BPH-811: had relatively low activity in these assays (IC_{50} values of 133 and



28 μM, respectively) and good activity in the *P. falciparum* growth inhibition assay (IC_{50} = 690 nM and 1 μM, for BPH-703 and BPH-811, respectively). Both compounds also have favorable logP values (logP ~2; Table 3.2) in addition to good cell activity, as illustrated in Figure 3.2E. Remarkably, of the 564 compounds assayed, the two most promising leads were both found to

both be simple derivatives of risedronate and zoledronate in which the 1-OH group is replaced by an H, and the aryl group contains a C_{10} or C_{12}, hydrophobic substituent that improves logP and is expected to reduce bone binding, as shown previously with BPH-715 (19).

Target Identification.

The results presented above are of considerable interest since we find that two of the most potent inhibitors with low toxicity are both "lipophilic" analogs of the well-known bisphosphonate drugs zoledronate (Zometa®/Reclast®) and risedronate (Actonel®/Atelvia®) which target FPPS (3). This strongly suggested the possibility that both of these lipophilic bisphosphonates (LBPs) target one or more prenyl synthases. In the case of BPH-703, as illustrated in Figure 3.2F, it can be seen that this molecule can be thought of as being derived from a "fusion" of the key structural features found in zoledronate (the bisphosphonate and imidazolium/carbocation-isostere) with the hydrophilic diphosphate and lipophilic features found in FPP, or GGPP. In previous work, we found that the lipophilic pyridinium bisphosphonate BPH-715 was a potent inhibitor of both liver stage as well as blood state malaria parasites (14), but BPH-703 is far less active than is BPH-715 against three human cell lines (~133 μM vs. ~168 nM, for BPH-715), Table 3.2. In the case of BPH-811, it can again be seen (Figure 3.2F) that this compound is basically a more lipophilic analog of risedronate in which there is a decyloxy side-chain addition, and the 1-OH is replaced by H. The logP values for risedronate and zoledronate are -2.9 and -3.5, while those for the LBPs are much more positive (Table 3.2): BPH-703 = 2.5; BPH-811 = 1.5, which can be expected to result in enhanced cell activity (17).

To determine whether *Plasmodium* GGPPS is in fact a major target for the lipophilic bisphosphonates, we expressed GGPPS from *P. vivax* (whose x-ray structure is known and which has 73% identity and 89% similarity to the *P. falciparum* protein) and determined dose-

response curves for both inhibitors from which we find IC_{50} values of 1.2 μM (BPH-703) and 2.5 μM (BPH-811), corresponding to K_i values of 270 and 566 nM, consistent with a GGPPS target in *P. falciparum*. To further test this idea, we next carried out cell growth inhibition "rescue" experiments, as reported previously for *T. brucei* (34) and human tumor cell lines (20), in which we added either farnesol, geranylgeraniol or solanesol (C_{45}) to the growth medium. The results (Figure 3.10) clearly indicated that only GGOH (at 20μM) rescued cell growth, supporting a GGPPS target for both BPH-703 and BPH-811. These results are of interest since the extremely potent *Plasmodium* GGPPS inhibitors risedronate and zoledronate (16) have little activity in *P. falciparum* growth assays (4), and were not found to be hits in our initial screen, even though they have GGPPS IC_{50} values of ~500 nM (Table 3.2 and Ref.(16)) against *P. vivax* GGPPS. We thus next sought to determine how BPH-703 and BPH-811 bound to *Plasmodium vivax* GGPPS, whose x-ray crystallographic structure we recently reported (16) .

Crystallographic structures of *Pv*GGPPS-inhibitor complexes.

As we and others reported previously (30, 35, 36), there are four main ligand (substrate or inhibitor) binding sites in human and *S. cerevisae* GGPPS, illustrated in Fig. 2.3 (from Fig. 3 in ref (36). Two are polar binding sites where the diphosphates in the substrates (IPP, and FPP) bind, and two are hydrophobic sites, where the C_{15} (FPP) substrate or C_{20} (GGPP) products can bind. Bisphosphonate inhibitors can bind with their phosphonate groups in either polar binding site, and there are three district polar/non-polar motifs for bisphosphonates bound to *S. cerevisae* GGPPS (30, 36), as shown in Figure 3.3A, while in PvGGPPS, small bisphosphonates bind to the FPP site while the GGPP product spans the I***PP-F***PP site (Figure 3.3A). We thus next obtained the single crystal x-ray crystallographic structures of BPH-703 and BPH-811 bound to PvGGPPS, using co-crystallization. Full crystallographic and structure refinement details are

given in Table 3.1; crystallization details were as described previously (16). With PvGGPPS (PDB ID code 3LDW), zoledronate binds to the FPP site (in yellow), Figure 3.3A, B, and unlike the situation found with human and *S. cerevisae* GGPPS, is a potent inhibitor (IC_{50} ~510 nM, versus ~100 μM in human GGPPS (Table 3.2), due to the fact that the zoledronate bisphosphonate group binds to 3 Mg^{2+} in the PvGGPPS protein, just as in human FPPS (37, 38). With the zoledronate analog BPH-703, we find that the bisphosphonate headgroup also binds to 3 Mg^{2+} in the active site, just as the parent molecule zoledronate does, Figure 3.3A, B. In addition, the imidazolium group is situated in essentially the same position as found with zoledronate bound to PvGGPPS (PDB ID code 3LDW): a 0.79 Å rmsd based on a protein structure alignment; 0.37 Å for the imidazole, Mg^{2+} and bisphosphonate P, O atoms, based on a ligand/Mg^{2+} alignment. The polar headgroup in the lipophilic bisphosphonate thus binds in a very similar manner as does its zoledronate parent. However, the alkyl chain present in the lipophilic species BPH-703 occupies the FPP substrate hydrophobic site –not either or the GGPP product (or inhibitor) sites seen in other structures (30, 35, 36). The GGPP product site found in the yeast and human enzymes is blocked by R50 and the site that houses the "tail" of the FPP-analog inhibitor, *S-thiolo*-FPP, is blocked by F88. In addition, the BPH-703 side-chain can be seen to occupy a 13Å long hydrophobic pocket lined by numberous polar residues that spans both molecules in the dimer, Figure 3.11.

In the case of the lipophilic risedronate-analog, BPH-811, we again find that the bisphosphonate headgroup binds to a $[Mg^{2+}]_3$ cluster, with its long alkyl chain located in the FPP substrate side-chain site, Figure 3.4A. Protein contacts are shown in Figure 3.4B. The local (risedronate/Mg^{2+}) structure is very similar to that seen with risedronate bound to *T. cruzi* FPPS a (0.82 Å rmsd, protein alignment; 0.50 Å rmsd, ligand (Mg^{2+} local alignment, as described above),

as shown in Figure 3.4A. Moreover, the BPH-811 structure is very similar to that seen with BPH-703, is shown in Figure 3.4C. Overall, both PvGGPPS structures containing lipophilic bisphosphonates are thus characterized binding to a $[Mg^{2+}]_3$ cluster; close register of the risedronate, zoledronate cationic features in the liphophilic bisphosphonates with their more hydrophilic parent compounds; hydrophobic/alkyl groups that occupy the FPP side-chain pocket; and close register between the hydrophobic side-chains in both PvGGPPS inhibitors (an 1.19 Å rmsd for 10 carbons).

Interestingly, when both lipophilic bisphosphonates bind to PvGGPPS these do so by opening up a "hydrophobic tunnel" that spans the dimer structure. In the case of zoledronate (Figure 3.11A), this tunnel is closed but in the presence of either BPH-703 (Figure 3.11B) or BPG-811 (Figure 3.11C) it is readily apparent that there is a continuous tunnel that encompasses both GGPPS molecules, the surface shown being visualized using the ligand-free structures using the Pymol program (Delano Scientific, Palo Alto, CA), the ligands being superposed. In contrast, the polar part of GGPP binds to the IPP diphosphate-binding site while the side-chain is bent and occupies the FPP side-chain site, Figure 3.11D and the channel linking the two monomers is absent, as expected.

Isothermal Titration Calorimetry: Enzyme inhibition and Bone Binding.

Next, we investigated the thermodynamics of binding of BPH-703, BPH-811, as well as zoledronate and risedronate, to PvGGPPS by using isothermal titration calorimetry (ITC). The headgroups in the lipophilic bisphosphonates bind to PvGGPPS in a s similar manner to that seen with risedronate and zoledronate, but how does alkyl substitution affect the thermodynamics of binding?

In earlier work, we investigated the thermodynamics of binding of six bisphosphonates to FPPS from *T. brucei* (39), and related results have been reported for bisphosphonates binding to human FPPS (37, 38). With bisphosphonates having cationic side-chains (e.g. pamidronate, ibandronate and pyridinium-1-yl bisphosphonates), binding was overwhelmingly entropy driven (39), due to the hydrophobic effect. However, with neutral side-chain containing species e.g. risedronate and deaza-risedronate, binding was enthalpy driven. With PvGGPPS, binding is neither overwhelmingly entropy nor enthalpy driven. For BPH-703, the zoledronate analog, we find (Table S3) ΔG = -8.7 kcal mole^{-1} with ΔH = -5 kcal mole^{-1} and ΔS = 12 cal deg^{-1} mole^{-1}, Figure 3.5A, to be compared with ΔG = -9.5 kcal mole^{-1}, ΔH = -4.8 kcal mole^{-1} and ΔS = 16 cal deg^{-1} mole^{-1} for zoledronate, Figure 3.5B. So in this case, alkyl substitution (and loss of the 1-OH group) results in only a small overall change in ΔG of ~0.8 kcal mole^{-1}. This is an encouraging result since it indicates that making a major structural change in the inhibitor – adding a C_{12} side-chain, has only a relatively minor effect on GGPP binding, consistent with the enzyme inhibition results (Table 3.2,where of course Mg^{2+} and IPP-binding are also involved). This is in sharp contrast to the observation that BPH-703 is a potent inhibitor of *P. falciparum* growth *in vitro*, while the zoledronate parent has only very low activity where we find an IC_{50} = 167 μM (13). The same trend is seen with BPH-811 and its parent, risedronate, Figs. SA,B. With BPH-811, we find ΔG = -9.3 kcal mole^{-1}, ΔH = -2.9 kcal mole^{-1} and ΔS = 21 cal deg^{-1} mole^{-1} (Table 3.3) while for the risedronate parent, ΔG = -9.3 kcal mole^{-1}, ΔH = -2.5 kcal mole^{-1} and ΔS = 25 cal deg^{-1} mole^{-1}. In this case, the effect of adding the C_{10} decyloxy side-chain and remaining the 1-OH group does not affect ΔG. So: in both cases, conversion of the parent bisphosphonate to the more lipophilic analog, expected to have enhanced cell/tissue permeability

(17) and poor bone-binding capacity (18), has little or no effect on ligand binding and only a small effect on enzyme inhibition – even though there are major effects in cell-based assays.

It is also of interest to see how these results with GGPPS compare with those we reported previously, on FPPS (39). As can be seen in Figure 3.5C, the four ΔH, ΔS values fall on the line discussed previously for 6 bisphopshonates (at 2 pH values) binding to FPPS from *T. brucei* (39). The apparent "enthalpy-entropy compensation" is due simply to the fact that the overall range in ΔG is small while that for ΔH (and ΔS) is large so that $\Delta H = T\Delta S + b$ with a slope $\partial(\Delta H)_i/\partial(\Delta S)_i = T$, and as expected the slope found in Figure 3.5C is 298K, the temperature at which the ITC experiments were performed. What is surprising about these results is that the zoledronate/BPH-703 ΔS/ΔH results are very close, as are the risedronate/BPH-811 results, Figure 3.5C. Moreover, the two sets of results are themselves quite close, with neither the extreme entropy or enthalpy-driven behavior seen with FPPS (39). In FPPS, binding of all bisphosphonates containing a cationic side-chain in entropy driven (top right, Figure 3.5C). In GGPPS, however, we see behavior that is "in-between" these extremes. While this could of course be due to an inappropriate comparison between the different structures, the fact that the FPPS/GGPPS structures with e.g. zoledronate bound are essentially superimposable (16) suggests either that the charge center does not contribute to binding in GGPPS, or that the presence of the large neutral side-chain simply reduces the effects of the cation center. The former view is supported by the observation in *S. cerevisae* GGPPS that the cation center is not essential for GGPPS inhibition (36) unlike the situation with FPPS (3).

***In vivo* Testing Results.**

To see to what extend the lipophilic bisphosphonates have *in vivo* activity, we used a *P.chabaudi* suppressive test. We first injected mice with 10^6 parasites followed by 10 mg/kg of each

bisphosphonate, i.p., for 4 days. With BPH-703, there was a 100% reduction in parasitemia, Figure 3.6A, and a 100 % survival at day 14, Figure 3.6B, although there was weight loss that disappeared after treatment. In the case of BPH-811, there was an ~80% decrease in parasitemia and an 80% survival at day 14. We then carried out a second series of experiments using a reduced (3mg/kg, i.p.) dosing protocol. As with the 10 mg/kg dose, the best results were obtained with BPH-703 (Figure 3.6C,D) where there was a 100% survival (Figure 3.6D) and only ~ 8% parasitemia (Figure 3.6C), 14 days after dosing. As can be seen in Table 3.2, BPH-703 has good *in vitro* IC_{50} and logP values and little effects on human cell lines as well as good *in vivo* activity and thus represents a potentially attractive lead for further development.

3.4 Conclusions

Overall, the results presented here are of interest since they show that GGPPS is a promising target for anti-malarials, and that the lipophilic analogs of the bone resorption drugs zoledronate and risedronate have activity boh *in vitro* as well as *in vivo*. BPH-703 is a lipophilic analog of zoledronate in which the 1-OH group is removed (to prevent bone binding) and the ring is alkylated, to enhance lipophilicity. BPH-811 is an analog of risedronate in which, again, the 1-OH group is removed and a lipophilic tail added, to increase cell permeability, decrease bone-binding, and enhance tissue distribution. Both compounds have potent activity *in vitro* and *in vivo* while the parent drugs, zoledronate and risedronate, have only modest activity against *P. falciparum* in intra-erythrocytic assays, and risedronate, while reducing parasitemia, has no effect on survival (15). Also of interest is the observation that the lipophilic bisphosphonate inhibitors in the *Plasmodium* GGPPS structures adopt very similar binding poses to those of their parent compounds bound to both GGPPS as well as FPPS, and have similar enzyme activity and

ΔG values, for GGPPS binding. It thus seems likely that these and related lipophilic bisphosphonates will have improved activity over that already reported with more hydrophilic bisphosphonates (4-15) against other parasitic protozoa, such as *T. brucei*, *T. cruzi*, *Leishmania* spp., *T. gondii*, *C. parvum* and *E. histolytica*.

3.5 Materials and Methods

***P. falciparum* culture.**

P. falciparum strain 3D7 was maintained *in vitro* in human O^+ erythrocytes at 3% hematocrit in RPMI 1640 media supplemented with L-glutamine, 25 mM HEPES and sodium bicarbonate. Albumax I (0.5%) and HT Supplement (0.1 mM hypoxanthine and 16 μM thymidine) were added to complete the media.

***P. falciparum* drug screening and rescue experiments.**

A total of 564 compounds were screened at 10 μM, in duplicates. For screening, the *P. falciparum* culture was adjusted to 2% hematocrit, 0.5% parasitemia, dispensed by a WellMate (Thermo) to 384 wells plate (Greiner) containing the compounds (final volume 50 μL) and incubated for 72 hours. Chloroquine, artemisinin and DMSO were used within the assay plates to serve as controls.

After three days, the parasite lactate dehydrogenase (pLDH) assay was used to assess drug efficacy. Briefly, at the end of 72 h the plates were frozen overnight at -20°C. After thawing, the plates were shaken for 45 seconds at 1700 rpm in a Mix Mate (Eppendorf) and 5 μL of the lysate transferred into the corresponding well of another plate containing 30 μL of Malstat Reagent (40) and incubated for 2 h. The absorbance (650 nm) was read using a Spectramax M5 (Molecular Devices, USA). For hit confirmation, those compounds selected as hits were cherry-

picked, serially diluted (range 9.7 nM to 40 μM) and tested in triplicates under the same conditions as used in the initial screening. To confirm the activity of the compounds using another readout, we adapted a SybrGreen assay (41) to a 384 well plate format and tested a randomly chosen sample of the hits (15%), as follows. After incubation with the compounds, 25 μL of lysis buffer (30mM Tris pH 7.5, 7.5 mM EDTA, 0.012% saponin, 0.12% Triton X-100 plus 0.3 μL SybrGreen per mL) were added to each well. The plates were shaken for 45 seconds at 1700 rpm in a Mix Mate, incubated for 1 hour at room temperature, and fluorescence read (Ex 485nm/ Em 530nm) using a Spectramax M5.

For cell growth inhibition rescue experiments, selected compounds were tested over a range of concentrations (39 nM to 20 μM). Parasites were incubated with either the compound alone, the compound and 20 μM farnesol or the compound and 20 μM geranylgeraniol in 384 well plates and under the same conditions as the hits confirmation experiments.

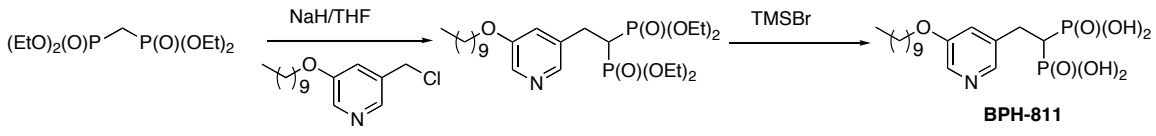
Cell Growth Inhibition Assays, Human FPPS Inhibition and Human GGPPS Inhibition.

The tumor cell, human FPPS and human GGPPS inhibition assays were carried out as described in our previous work (42).

Synthetic Aspects.

All reagents used were purchased from Aldrich and Alfa Aesar. The purities of all bisphosphonic acids were routinely monitored by using ^{1}H NMR and ^{31}P NMR spectroscopy on Varian (Palo Alto, CA) Unity spectrometers and by microchemical analysis. **BPH-688**, **BPH-710** and **BPH-715** were available from previous work (43).

2-(5-decyloxypyridin-3-yl)ethylidene-1,1-bisphosphonic acid (BPH-811).



BPH-811

To a suspension of NaH (480 mg, 60% in mineral oil, 12 mmol) in anhydrous THF (20 mL) at 0 °C was added tetraethyl methylenebisphosphonate (2.88 g, 10 mmol) and the mixture was allowed to stir at room temperature for 30 min. 3-Decyloxy-5-chloromethyl-pyridine (2.83 g, 10 mmol, in 5 mL THF) and NaI (1.8g, 12 mmol) were added, and the reaction mixture was refluxed under nitrogen for 2 h. The mixture was then queched with saturated aq. NH_4Cl at room temperature. Extraction with ethyl acetate, and flash chromatography (ethanol: ethyl acetate, 5:95) afforded the tetraethyl ester as a colorless oil (2.6 g, 50% yield). Dealkylation using 8 equivalent of TMSBr gave BPH-811 as a white powder (1.7 g, 80%). Anal. ($C_{17}H_{31}NO_7P_2 \bullet H_2O$). ^{1}H NMR (D2O, 400 MHz): δ 7.9 (1H, s), 7.8 (1H, s), 7.2 (1H, s), 4.0 (3H, t, *J* = 6.4Hz), 2.9 (2H, td, *J* = 15Hz, 6.8Hz), 1.9 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.5-1.6 (2H, m), 1.0-1.15 (12H, m), 0.63 (3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 19.6.

All pyridinium-1-yl bisphosphonates (**BPH-816, BPH-776, BPH-805, BPH-725, BPH-817, BPH-820, BPH-638**) were prepared as described previously (44).

2-(3-decylaminopyridium)-1-yl-ethylidene-1,1-bisphosphonic acid (BPH-816).

Anal. ($C_{17}H_{32}NO_6P_2 \bullet 0.75H_2O$). ^{1}H NMR (D2O, 400 MHz): δ 7.91 (1H, s), 7.84 (1H, d, J = 5.2Hz), 7.30-7.41 (2H, m), 4.5 (2H, t, *J* = 15Hz, 6.8Hz), 3.14 (2H, t, *J* = 6.8Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.5-1.6 (2H, m), 1.0-1.4 (14H, m), 0.67 (3H, t, *J* = 6.0 Hz). ^{31}P NMR (D_2O, 162 MHz): δ 14.6.

2-(3-undecyloxypyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (BPH-776).

Anal. ($C_{18}H_{31}NO_6P_2 \bullet 1.5H_2O$). ^{1}H NMR (D2O, 400 MHz): δ 8.49 (1H, s), 8.35(1H, d, J = 6Hz), 7.76 (1H, d, J = 9Hz), 7.6 (1H, d, J = 9Hz), 4.3 (2H, t, *J* = 15Hz, 6.8Hz), 4.0 (2H, t, *J* = 6.0Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.62-1.74 (2H, m), 1.0-1.3 (16H, m), 0.67(3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 14.6.

2-(3-Bromo-5-decyloxy-pyridium)-1-yl-ethylidene-1,1-bisphosphonic acid (BPH-805).

Anal. ($C_{17}H_{29}BrNNaO_7P_2$). ^{1}H NMR (D2O, 400 MHz): δ 8.65 (1H, s), 8.54 (1H, s), 8.03 (1H, s), 4.6 (2H, t, *J* = 15Hz, 6.8Hz), 4.1 (2H, t, *J* = 6.4Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.5-1.6 (2H, m), 0.9-1.4 (14H, m), 0.64(3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 14.3.

2-(3-Decylsulfanyl-pyridinium-1-yl)-1,1-bisphosphonic acid (BPH-817).

Anal. ($C_{17}H_{31}NO_6P_2S$•$0.1H_2O$) ^{1}H NMR (D2O, 400 MHz): δ 8.65 (1H, s), 8.50 (1H, d, J = 4.5Hz), 8.03 (1H, d, J = 8.5Hz), 7.67 (1H, t, J = 8.5Hz), 4.7 (2H, tt *J* = 15Hz, 6.8Hz), 2.97 (2H, t, *J* = 7Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.46-1.53 (2H, m), 1.0-1.4 (14H, m), 0.66(3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 14.5.

2-(3-Dodecylsulfanyl-pyridinium-1-yl)-1,1-bisphosphonic acid (BPH-820).

Anal. ($C_{19}H_{35}NO_6P_2S$) ^{1}H NMR (D2O, 400 MHz): δ 8.64 (1H, s), 8.51 (1H, d, J = 4.5Hz), 8.0 (1H, d, J = 8.5Hz), 7. 7 (1H, t, J = 8.5Hz), 4.7 (2H, m), 2.94 (2H, t, *J* = 7Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.46-1.53 (2H, m), 1.0-1.4 (18H, m), 0.64(3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 14.4.

2-(3-decylpyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (BPH-638).

Anal. ($C_{17}H_{31}NO_6P_2$•$0.5H_2O$). ^{1}H NMR (D2O, 400 MHz): δ 8.67 (1H, s), 8.64 (1H, d, J = 4.8Hz), 8.0 (1H, d, J = 8.0Hz), 7. 7 (1H, t, J = 8.0Hz), 4.7 (2H, tt, *J* = 15Hz, 6.8Hz), 2.67 (2H, t, *J* = 7.5Hz), 2.2 (1H, tt, *J* = 21 Hz, 6.8Hz), 1.46-1.53 (2H, m), 1.0-1.4 (14H, m), 0.68(3H, t, *J* = 6.0 Hz). 31P NMR (D_2O, 162 MHz): δ 14.8.

2-[3-(3,7-dimethyloctyloxy)pyridinium-1-yl]ethylidien-1,1-bisphosphonic acid (BPH-776).

Anal. ($C_{17}H_{29}NO_6P_2$•$1.75H_2O$). ^{1}H NMR (D2O, 400 MHz): δ 8.8 (1H, s), 8.42(1H, d, J = 6Hz), 7.82 (1H, d, J = 9Hz), 7.6 (1H, d, J = 9Hz), 4.8 (2H, m), 4.2 (2H, t, *J* = 6.0Hz), 2.2 (1H, tt, *J* = 21

Hz, 6.8Hz), 1.0-1.8 (10H, m), 1.0-1.3 (16H, m), 0.84 (2H, d, J = 6.5Hz), 0.79(3H, t, *J* = 6.5 Hz). 31P NMR (D_2O, 162 MHz): δ 14.6.

Protein expression and purification.

A clone encoding *P. vivax* GGPPS (PlasmoDB gene ID: Pv092040) with an N-terminally His_6-tagged fusion protein and a TEV protease site was expressed in *Escherichia coli* BL21-codon Plus (DE3) RIL (Stratagene) at 15 °C in baffled flasks. Cells were lysed by sonication in the presence of Benzonase Nuclease (Novagen), and a protease inhibitor cocktail (Sigma), and the protein then purified chromatographically by using a Ni-NTA resin. EDTA was added immediately to 1mM to the elution fraction and 5mM of DTT added after 15 minutes. The eluted protein was then concentrated and loaded onto a Sephadex 26/60 gel filtration column and fractions containing PvGGPPS collected for further experiments.

***in vitro* enzyme assay.**

The *P. vivax* GGPPS assays were carried out as describe in previous work (45).

Molecular Docking.

Docking calculations by performed by using PvGGPPS (PDB ID code 3RMB) minus the BPH-703 ligand as a template. The target protein was prepared using the protein preparation wizard in Maestro 8.0 (46). For monomer docking, only chain A was selected. For preparation of the ligand LigPrep (47) was used to geometry optimize the ligand at a target pH of 7.0. A grid was generated to cover the binding site, as well as the dimer interface in the case of dimer docking Glide (48) was used for docking calculation under, standard precision (SP) mode. The top 5 poses from the docking calculation was compared to the crystallographically-determined ligand pose by measuring the C^{α} RMSD with the superposition tool in Maestro.

Protein Crystallization.

Wild type PvGGPPS was eluted from column using 500 mM NaCl, 10 mM HEPES, pH 7.5 and concentrated to 15 mg/mL. Co-crystallization of PvGGPPS with the bisphosphonate inhibitors was carried out by first incubating PvGGPPS with 1mM bisphosphonate (BPH-703 or BPH-811) and 1mM $MgCl_2$ for 30 minutes, on ice. After incubation, 1.5 uL of crystallization buffer (19% PEG 3350, 1mM $MgCl_2$, 0.1M Tris-HCl pH7.4 and 0.1 M $LiSO_4$) was added to 1.5 uL of incubated protein by using the hanging-drop method, at 15˚C. Crystals appeared the next day. Data collection was carried out at [ADD] and indexed, integrated and scaled by using the HKL2000 program package (49). Structures were determined by using the molecular replacement method using the Phaser program (50). The PvGGPPS structure 2Q58, minus the ligand was used as a template. Further model building and refinement employed Coot (51), Refmac (52, 53), CNS (54), and the ProDRG server (55). Figures were obtained by using PyMol (56).

Isothermal Titration Calorimetry.

Isothermal titration calorimetry measurements were carried out at 298K using a MicroCal VP-ITC (MicroCal, Northampton, MA). 10 μL of 0.4 mM bisphosphonates were titrated from a 250-μL syringe (rotating at 270 rpm) into the sample cell containing 1.42 mL of a 0.03 mM PvGGPPS solution. The buffer solution was 50 mM Hepes (pH 7.4) and 1 mM $MgCl_2$. For running parameters, the duration of injection was set to 19.2 s, and the delay between injections was set as 300 s. The initial delay prior to the first injection was 60 s. To derive the heat associated with each injection, the area under each heat burst curve (microcalories per second versus seconds) was determined by integration (using Origin version 5.0 software; MicroCal, Northampton, MA). Fitting to a one-site binding model gave good accord with experiment.

***In vivo* experiments.**

Mice were infected by intraperitoneal (i.p.) injection of *Plasmodium chabaudi* AJ strain ($1x10^6$ parasites/mouse – day 0). Mice were separated into groups of 5 and compound treatments were given for 4 consecutive days (day 0-3) at 3, 10 (i.p.) or 100 mg/kg (gavage). Parasitaemia was monitored by using GIEMSA stained slides prepared with a sample of blood collected from the tail vein. Blood samples were collected daily for 10 consecutive days (days 3-12). Parasitaemia was determined by counting of at least 1000 cells in each sample (3). Mortality was followed for 30 days. All procedures were in accord with Korean regulations and were approved by Institut Pasteur-Korea Ethics Committee.

3.6 Tables and Figures

Table 3.1. Data collection, phasing, and refinement statistics.

Data collection	PvGGPPs-Mg-703	PvGGPPs-Mg-811
Space group	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)[a]	45.59-2.61 (2.67-2.61)	50.0-2.80 (2.85-2.80)
	Unit cell dimensions	
a (Å)	107.68	99.74
b (Å)	107.95	106.58
c (Å)	140.70	139.03
	No. of reflections	
Unique	50687	37133
Completeness (%)	99.15	98.60
R_{merge} (%)	8.3 (30.5)	14.1 (65.9)
I/σ(I)	28.9 (7.29)	14.0 (1.05)
	Refinement	
No. of reflections	44719(3340)	34572(2237)
R_{work} (%)	20.67(25.2)	20.69(25.0)
R_{free} (%)	24.75(31.2)	29.92(36.3)
	Geometry deviations	
Bond lengths (Å)	0.018	0.013
Bond angles (°)	1.7	1.5
No. of all protein atoms	12025	12074
Mean B-values (Å^2)	34.5	42.6
No. of all cofactor atoms		
Mean B-values (Å^2)	15.4	36.85
No. of water molecules	266	254
Mean B-values (Å^2)	29.2	50.6
Ramachandran plot (%)		
Most favored	96.5	92.3
Additionally allowed	2.77	5.61
Generously allowed	0.78	1.42

Table 3.2. Enzyme and cell growth inhibition results.

compounds	LogP	*P. falciparum in vitro*	PvFPPS IC_{50}	PvGGPPS IC_{50}	Cancer Cell Average	HsGGPPS IC_{50}	HsFPPS IC_{50}
BPH-2 (risedronate)	-2.5	129	0.59	0.5	191	350	0.24
BPH-91 (zoledronate)	-2.9	167	0.79	0.51	15.8	100	0.11
BPH-638	2.4	0.59	3.8	0.78	0.16	0.4	1.57
BPH-688	1.4	0.73	0.57	1	2.3	1.23	0.26
BPH-703	2.5	0.69	1.8	1.2	133	3.3	1.7
BPH-710	1.4	0.79	0.62	1.1	0.73	2.1	0.009
BPH-715	2.1	0.8	1.9	NA	0.17	0.25	0.032
BPH-728	1.9	0.72	0.3	0.74	7.5	15	1.2
BPH-776	2.6	0.73	1.9	1.3	0.04	2.2	1.16
BPH-805	2.9	0.75	0.87	1.4	3.1	0.51	120
BPH-811	1.5	0.55	5.2	2.5	28	4.5	1.31
BPH-816	1.8	0.47	0.73	0.66	0.32	1.3	0.83

Table 3.3. ITC results.

compounds	ΔG	ΔH	ΔS
BPH-2 (risedronate)	-9.32	-2.47	23.0
BPH-91 (zoledronate)	-9.54	-4.83	15.8
BPH-703	-8.69	-5.02	12.3
BPH-811	-9.30	-2.92	21.4

Figure 3.1. Isoprenoid biosynthesis *Plasmodium* spp. and potential targets. Bisphosphonates inhibits FPPS/GGPPS as well as the long chain prenyl synthases involved in quinone biosynthesis, protein prenylation and dolichol formation; phosphonosulfonates and realated species inhibit head-to-tail prenyl transferases such as squalene and dehydrisqualene synthase and may target phytoene synthase; cationic nitrogen-containing species such as Ro48-8071 inhibit diverse prenyl transferases.

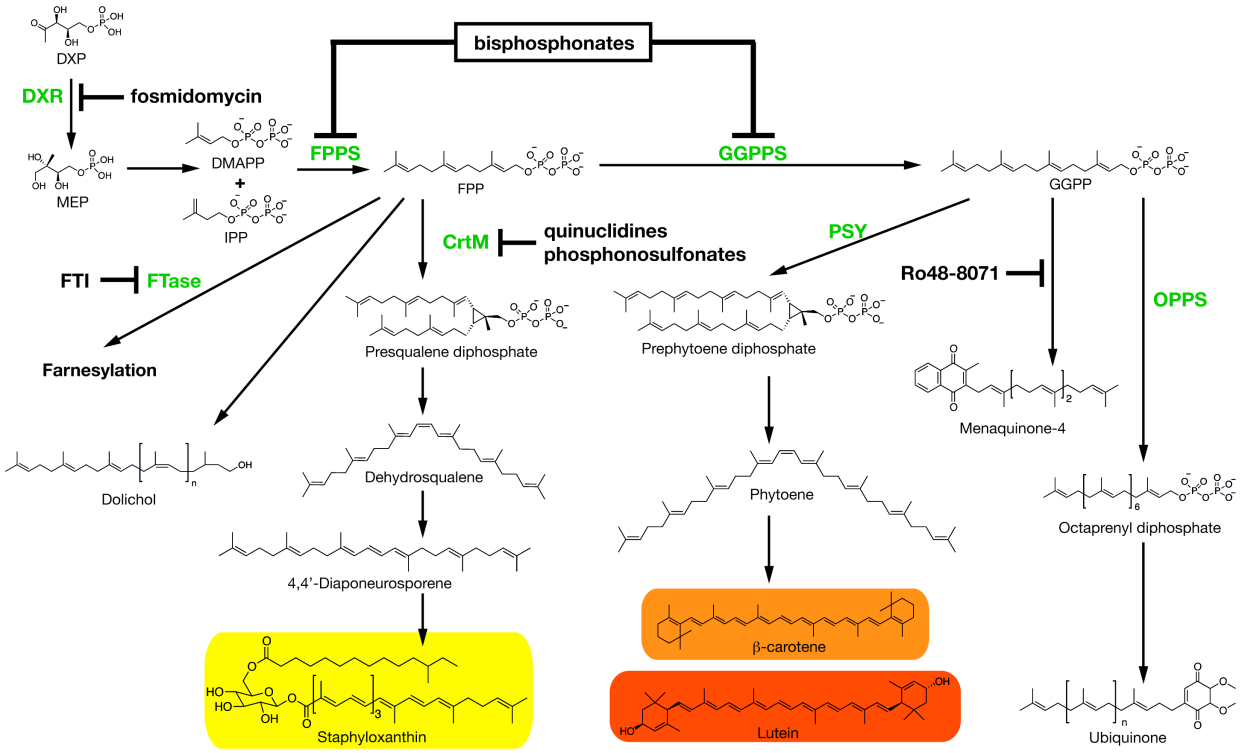


Figure 3.2. Selection of lead compounds for furhter investigation. (**A**) Representative screening results. Black: negative control (PBS; N=399 wells); blue: positive control (artemisinin. 192nM; N=181); gree, test compounds (N=564 compounds, 10μM). (**B**) Positive and negative controls along, Z′ = 0.72. (**C**) Dose-response curves for humen cell growth inhibition in NCIH-460, MCF-7 and SF-268 cell lines. Red: BPH-703; blue: BPH-776. (**D**)Therapeutic index plot for hits: BPH-703 and BPH-811 had good activity and low toxicity. (**E**) Scatter plot of *P. falciparum* pIC_{50} (=-$\log_{10}IC_{50}$, M) versus logP values indicating good pIC_{50}, logP values for compounds selected from (**D**). (**F**) Schematic illustration illustrating combination of structureal features of bisphosphonates and FPP substrate to yield lipophilic bisphosphonates.

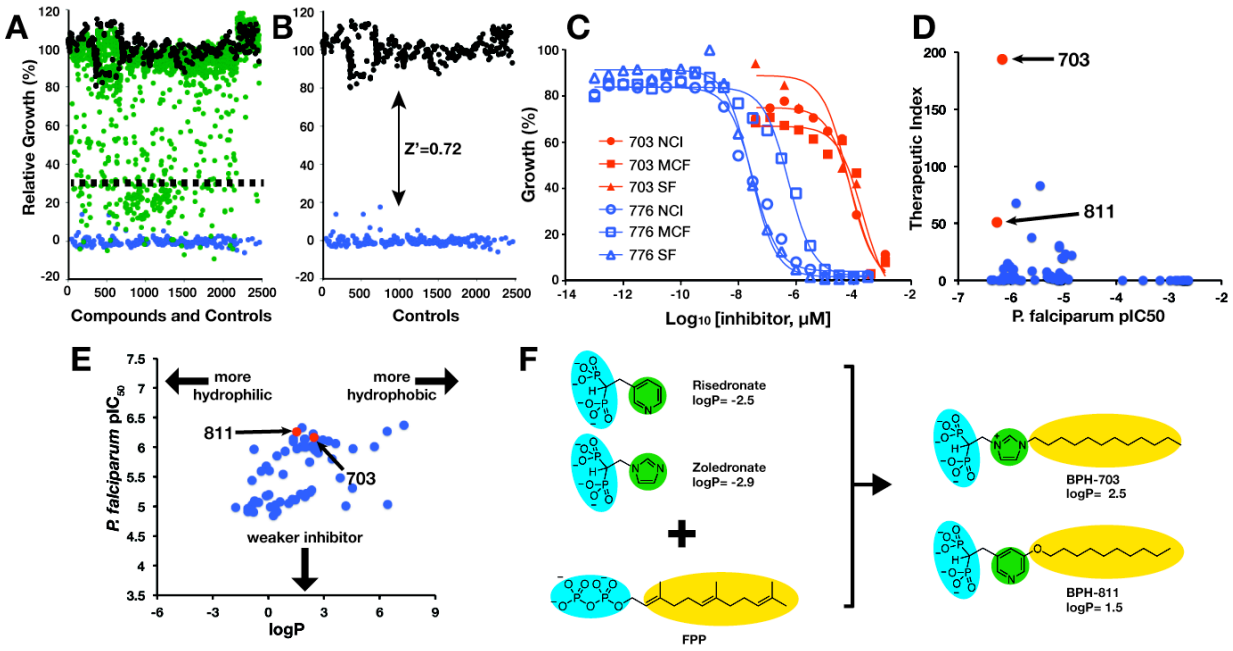


Figure 3.3. Crystallographic structure of the lipophilic zoledronate analog BPH-703 bound to PvGGPPS. (**A**)-(**D**) crystallographic structures of IPP, FPP, GGPP and several bisphosphonates bound to human, yeast or *P. vivax* GGPPS, after Fig.4. in (36). (**A**) the GGPP binding site in human and yeast (pH5.5) GGPPS. (**B**) the *P. vivax* zoledronate, *T. cruzi* risedronate and yeast FPP site. On both A, B, the ligand polar groups bind to the Mg^{2+}/Asp site, top left. (**C**) the I***PP-GG***PP site occupied by a bisphosphonate inhibitors. (**D**) the GGPP product site occupied in yeast GGPPS (pH=7.5) or PvGGPPS. (**E**) Superposition of BPH-703 bound to PvGGPPS with bound zoledronate. This is the FPP-binding site seen in (**B**) with the yeast enzyme, but the distal end of the inhibitor occupying a hydrophoic tunnel that spans the two monomers in the GGPPS dimer. (**F**) Protein-ligand interactions for BPH-703. (structures are rendered using PyMOL from Delano Scientific, Palo Alto, CA).

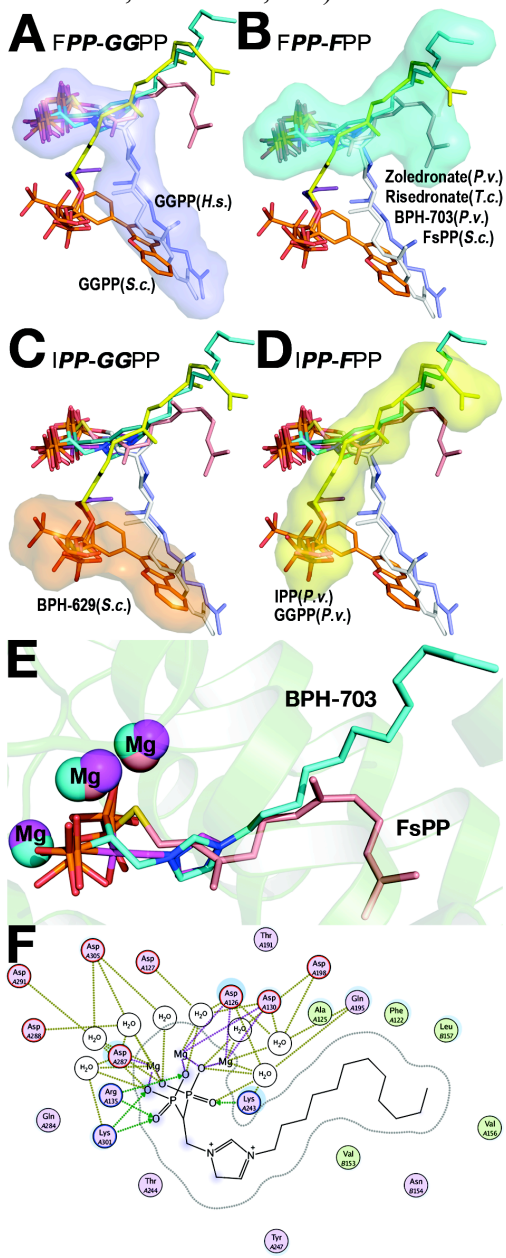


Figure 3.4. Crystallographic struture of the lipophilic analog of risedronate and BPH-811. (**A**) BPH-811 superposed on the risedronate-TcFPPS structure (**B**) protein-ligand interaction for BPH-811. (C) Overlay of BPH-811 structure with BPH-811 structure. Both occupy the hydrophobic tunnel between the dimer molecules.

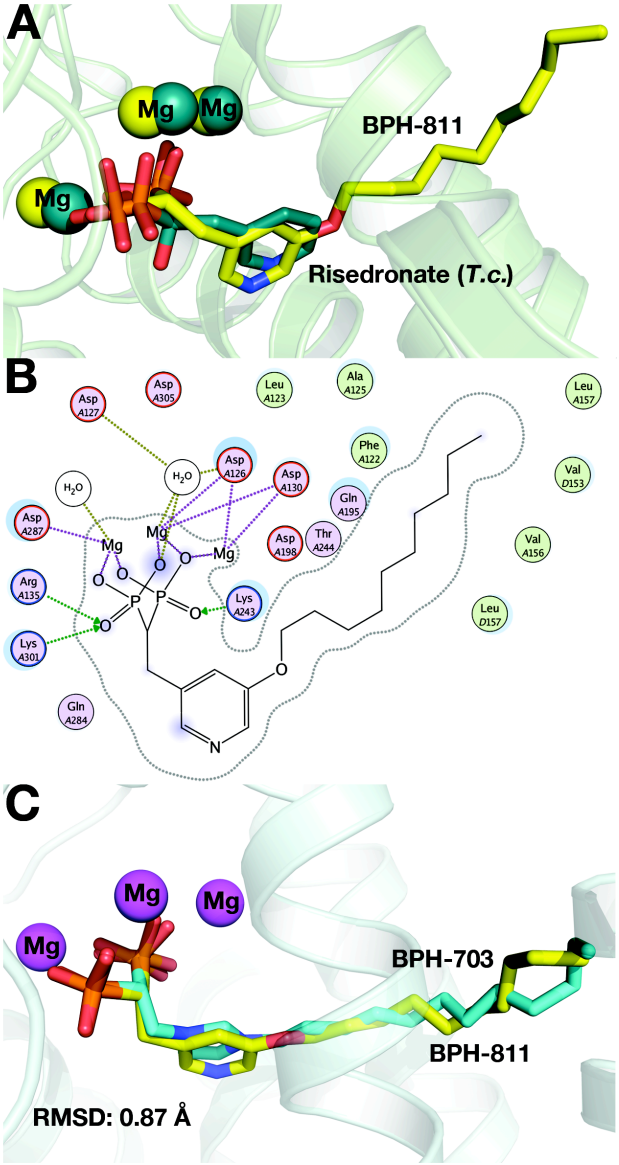


Figure 3.5. Isothermal titration calorimetry results. (**A**) zoledronate binding to PvGGPPS. (**B**) BPH-703 binding to PvGGPPS. (**C**) **ΔH/ΔS** plot for zoledronate, BPH-703, risedronate and BPH-811 binding to PvGGPPS (in red) superposed on risedronate, zoledronate and other bisphosphonates binding to *T. brucei* FPPS. (the latter based on result in (39)). There is no evidence for the dominance of entropy driven binding seen with cationic species binding to FPS, consistant with the obsevation that the presence of a cationic species is not important for GGPPS inhibitions.

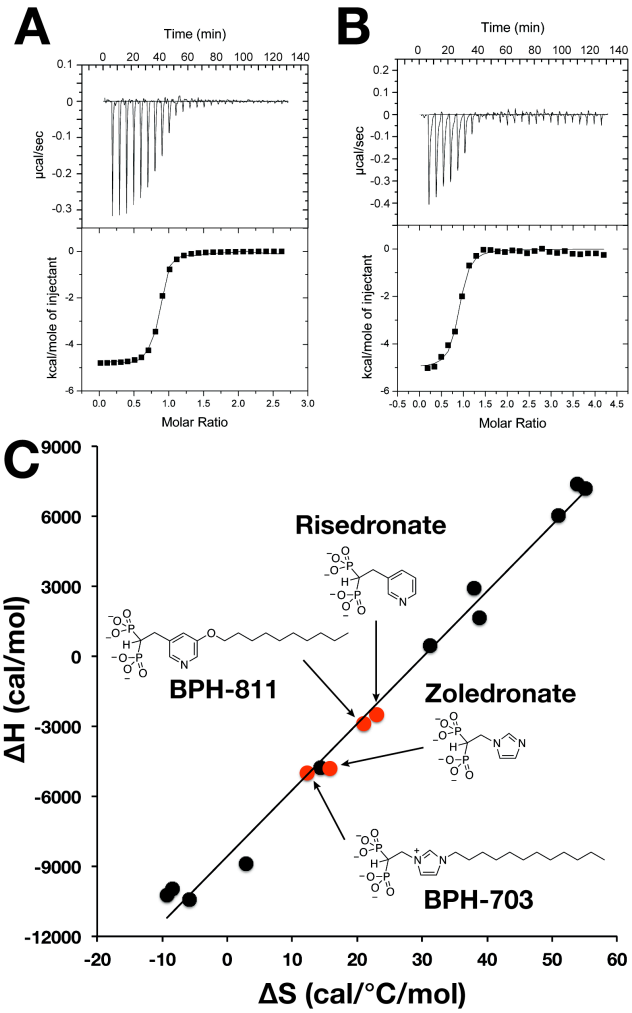


Figure 3.6. Effect of BPH-703 and BPH-811 on mouse survival and parasitemina. (**A**) Survival at 10mg/kg. (**B**) Parasitemia at 10mg/kg. (**C**) Survival at 3mg/kg. (**D**) Parasitemia at 3mg/kg. Black circle: PBS control; red triangle, BPH-703; blue triangle, BPH-811; green triangle, chloroquinine positive control.

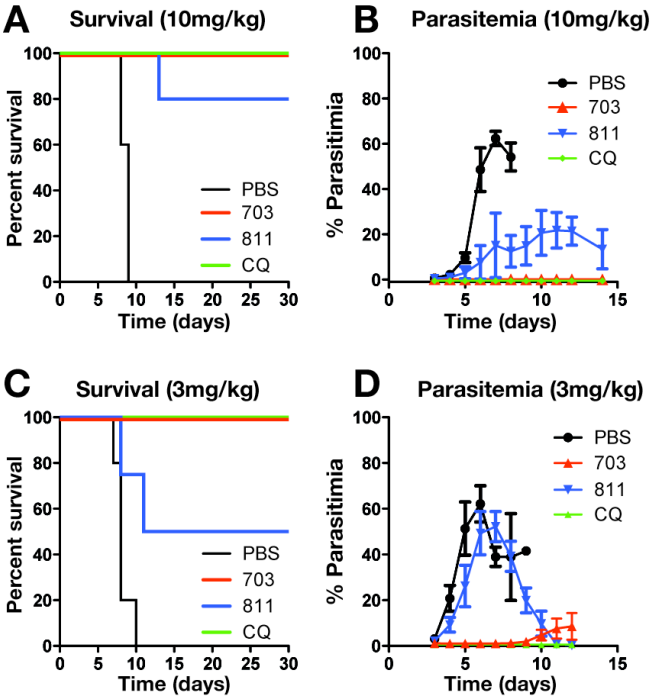


Figure 3.7. Plasmodium screening hits.

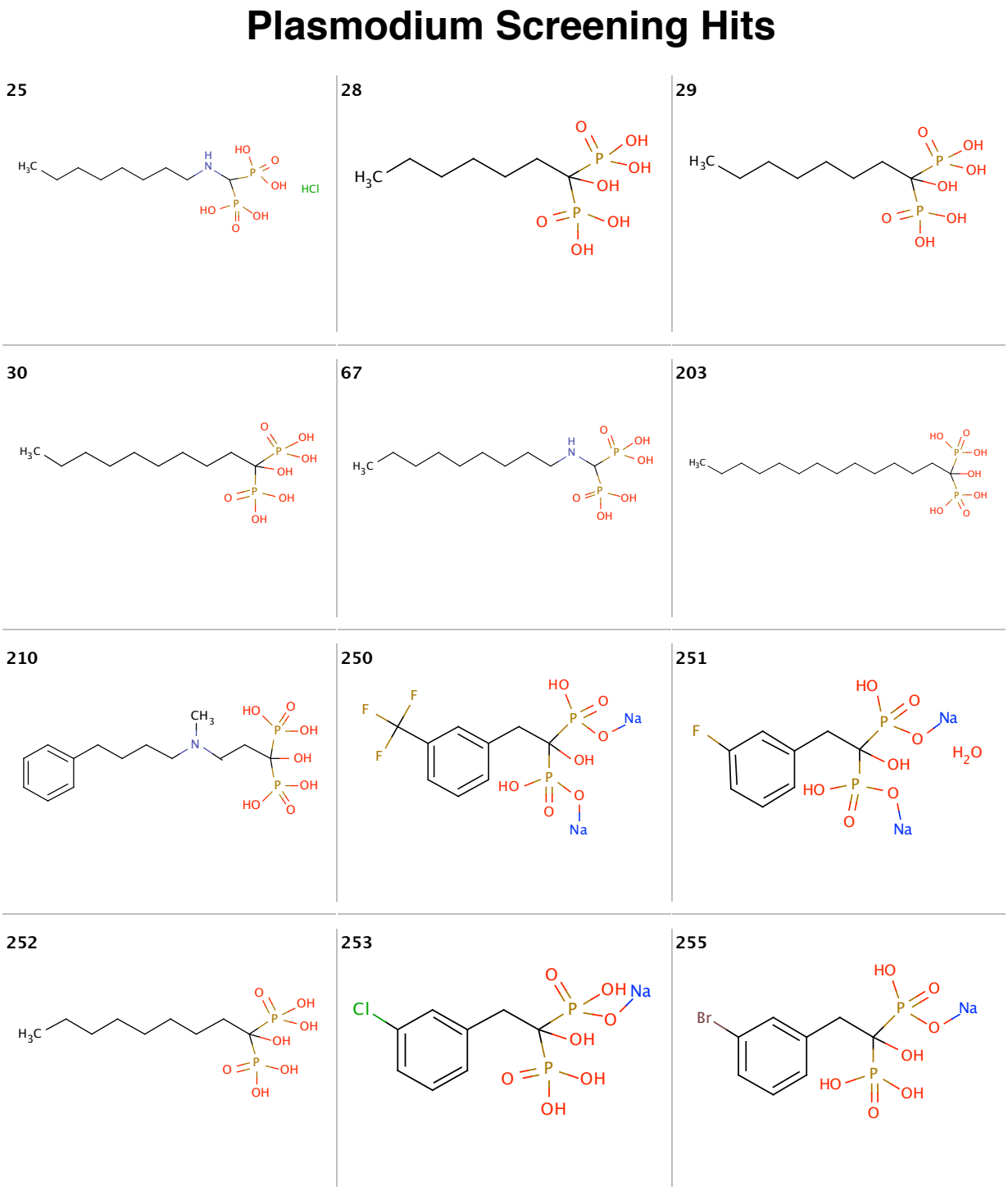


Figure 3.7. continued.

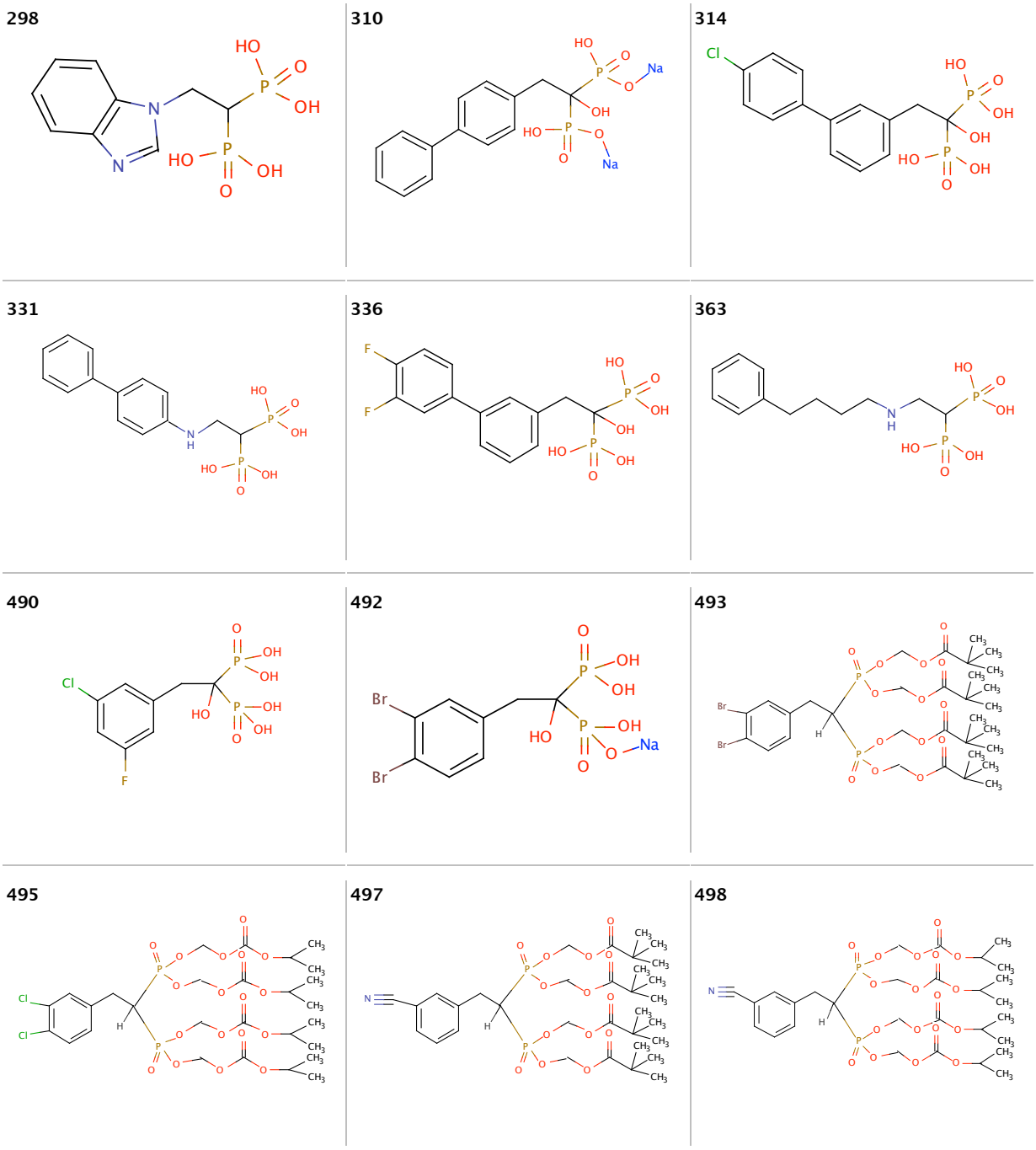
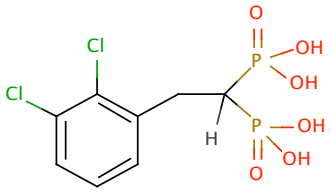
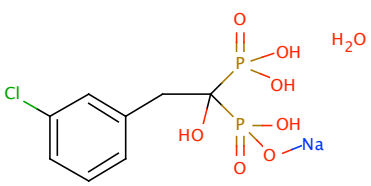


Figure 3.7. continued.

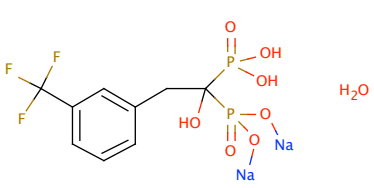
504



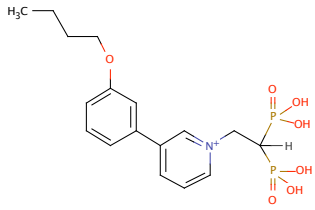
507



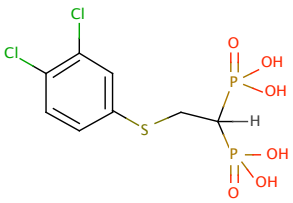
508



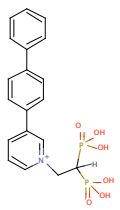
579



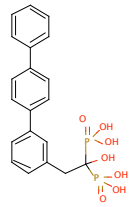
617



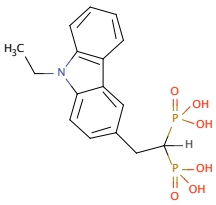
620



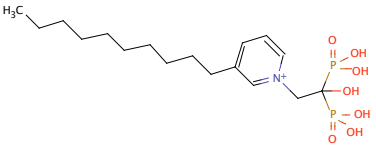
628



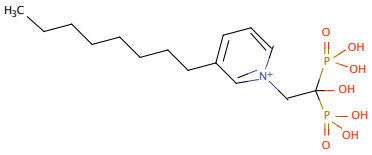
635



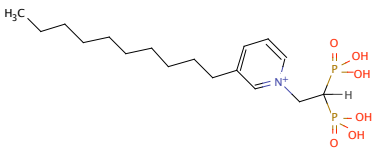
636



637



638



647

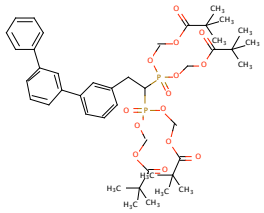


Figure 3.7. continued.

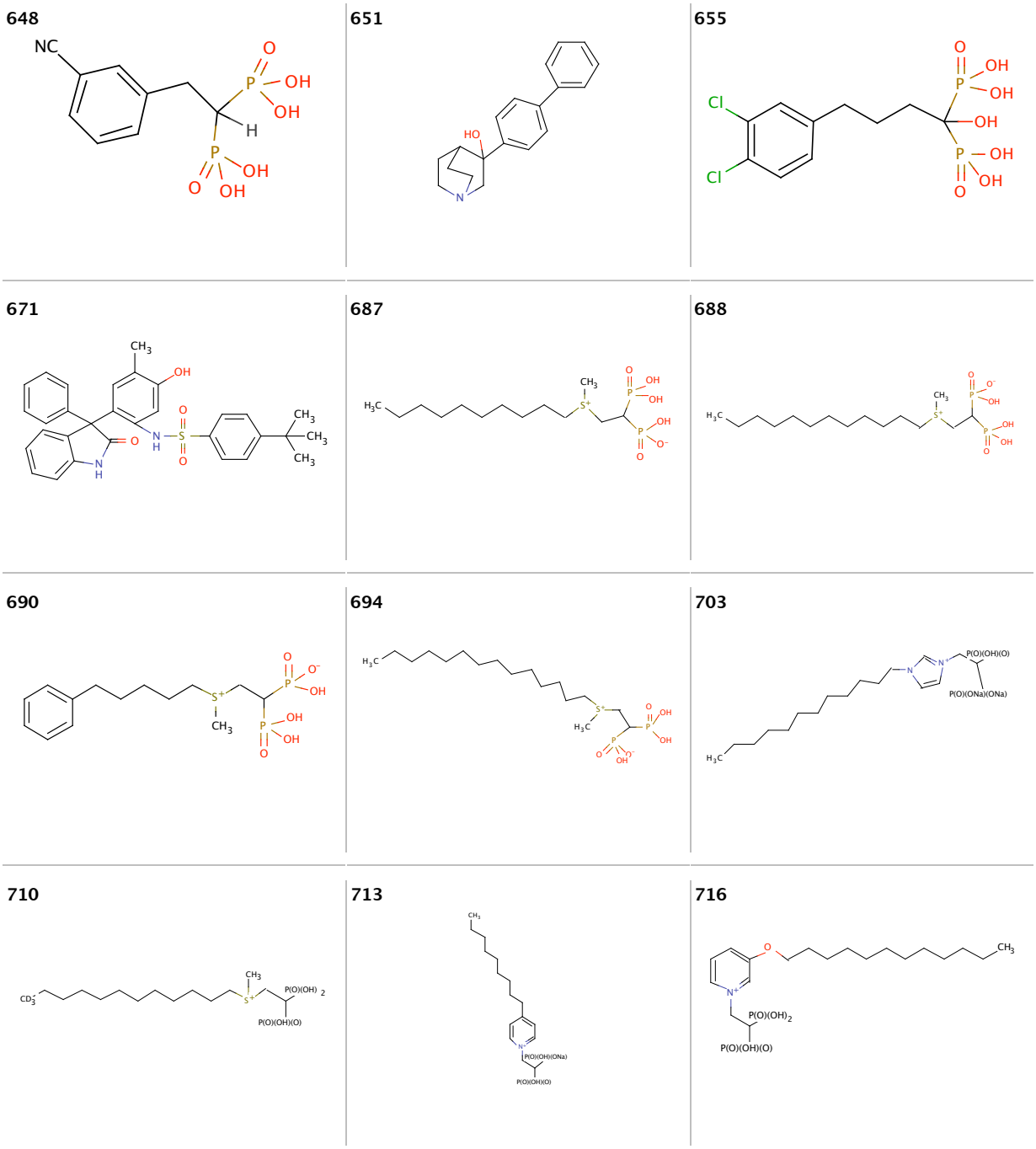


Figure 3.7. continued.

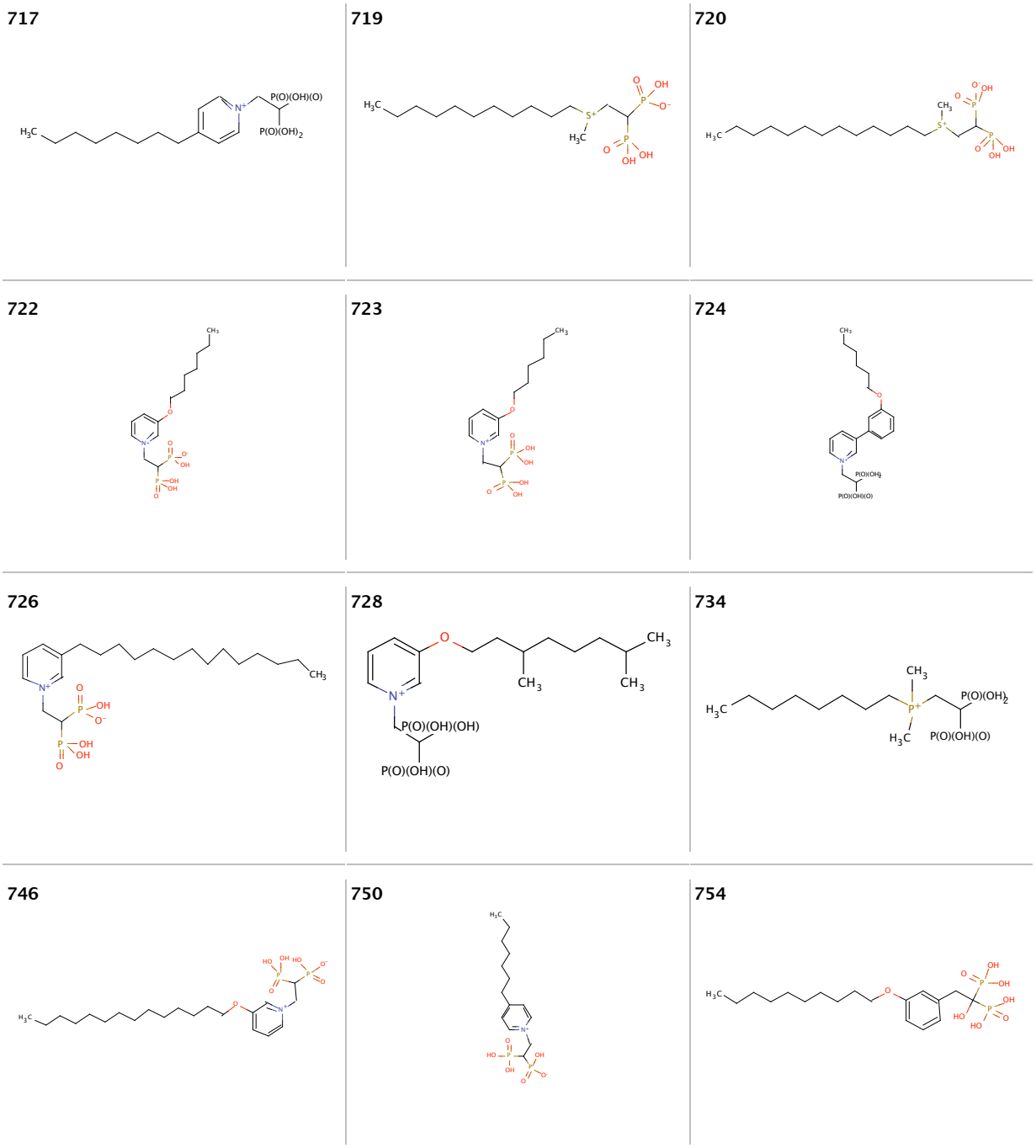


Figure 3.7. continued.

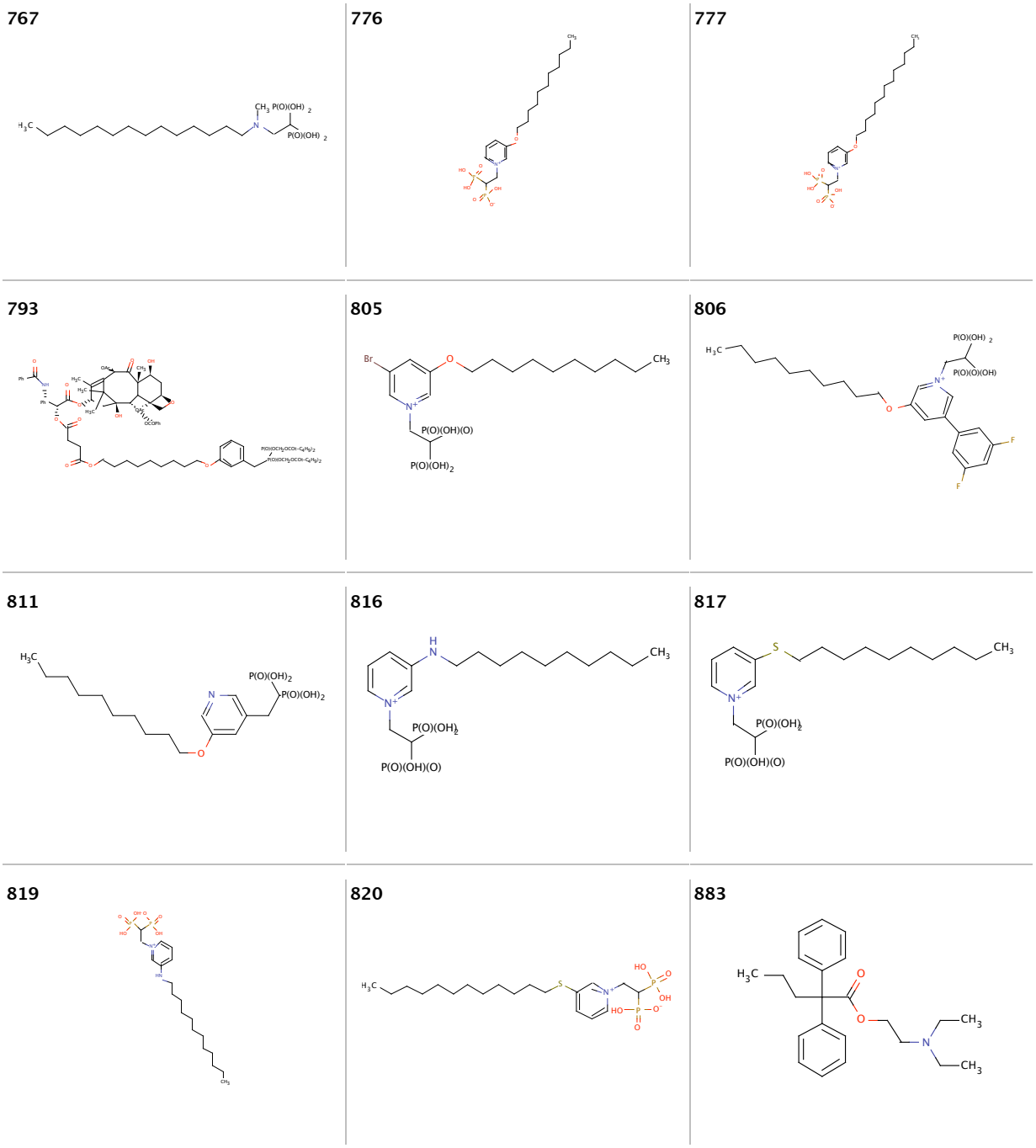


Figure 3.7. continued.

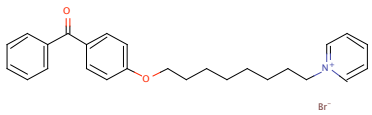
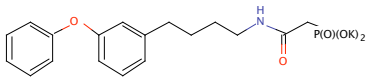
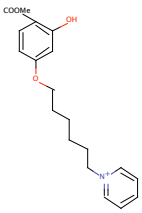
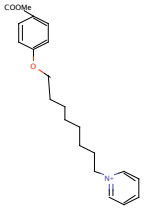
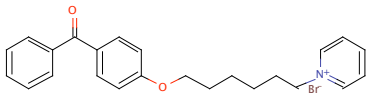
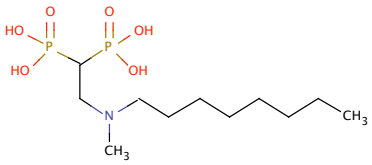
890 	**898** 	**907** 
908 	**921** 	**942** 

Figure 3.8. Dose response curve for hits.

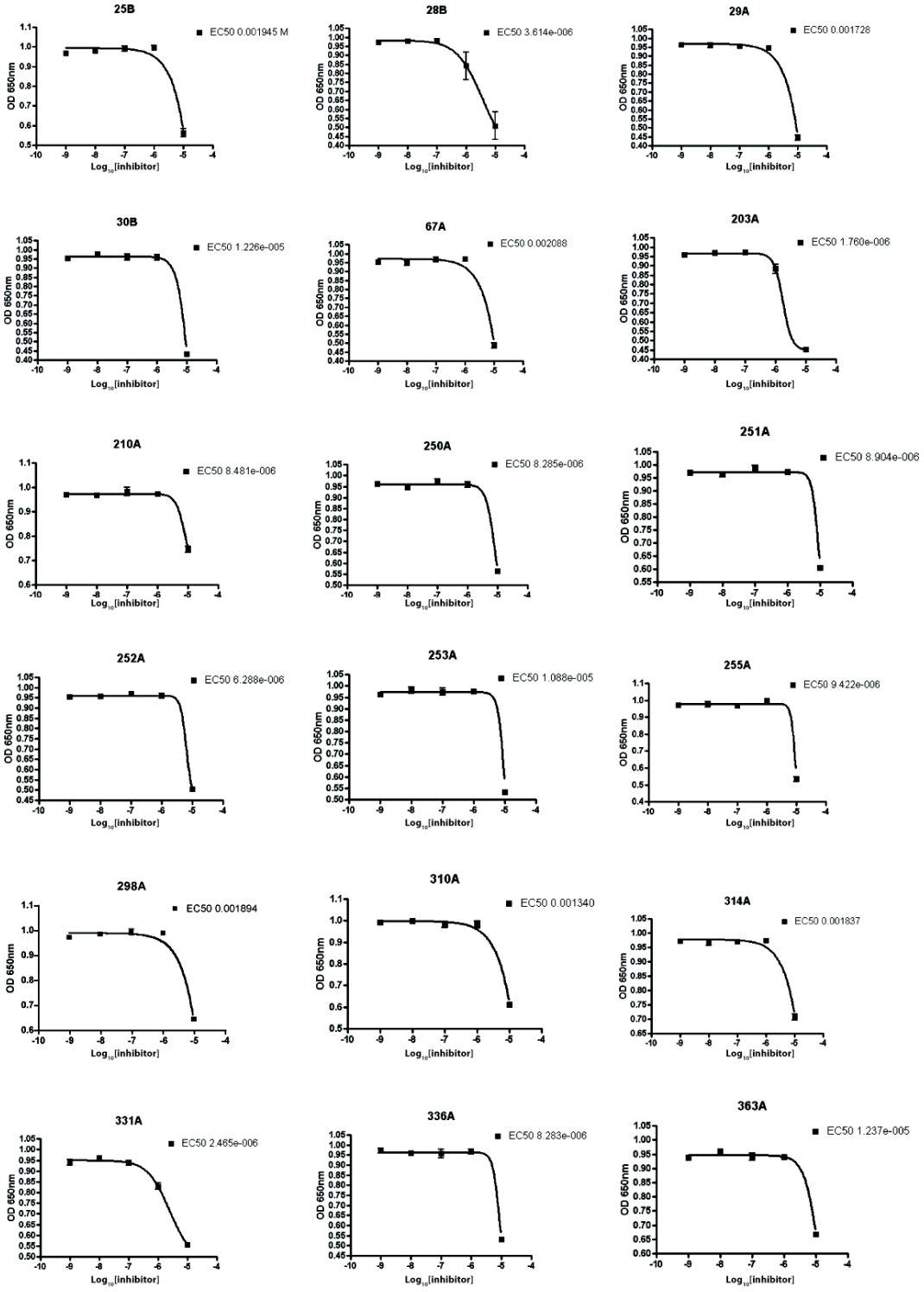


Figure 3.8. continued.

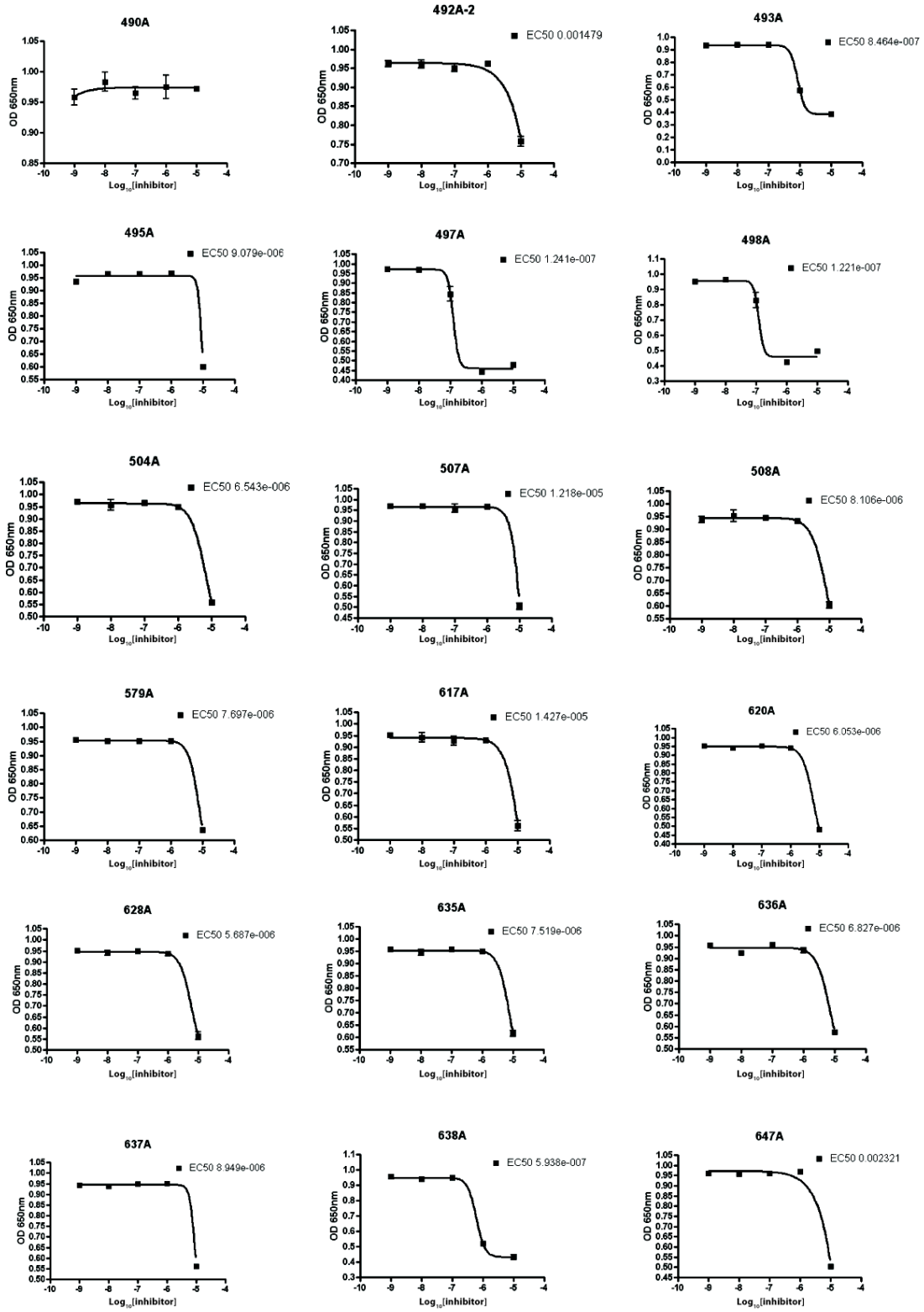


Figure 3.8. continued.

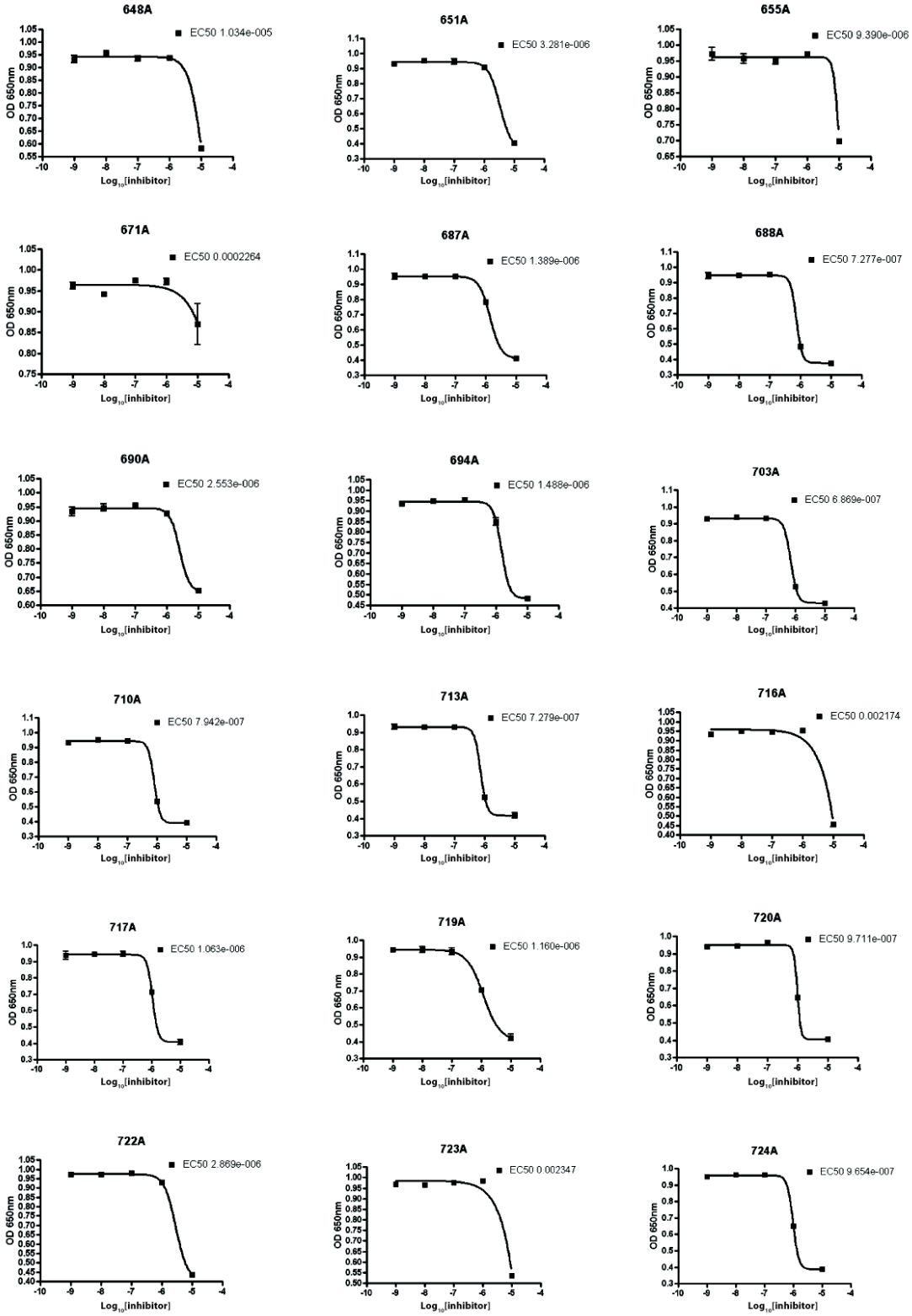


Figure 3.8. continued.

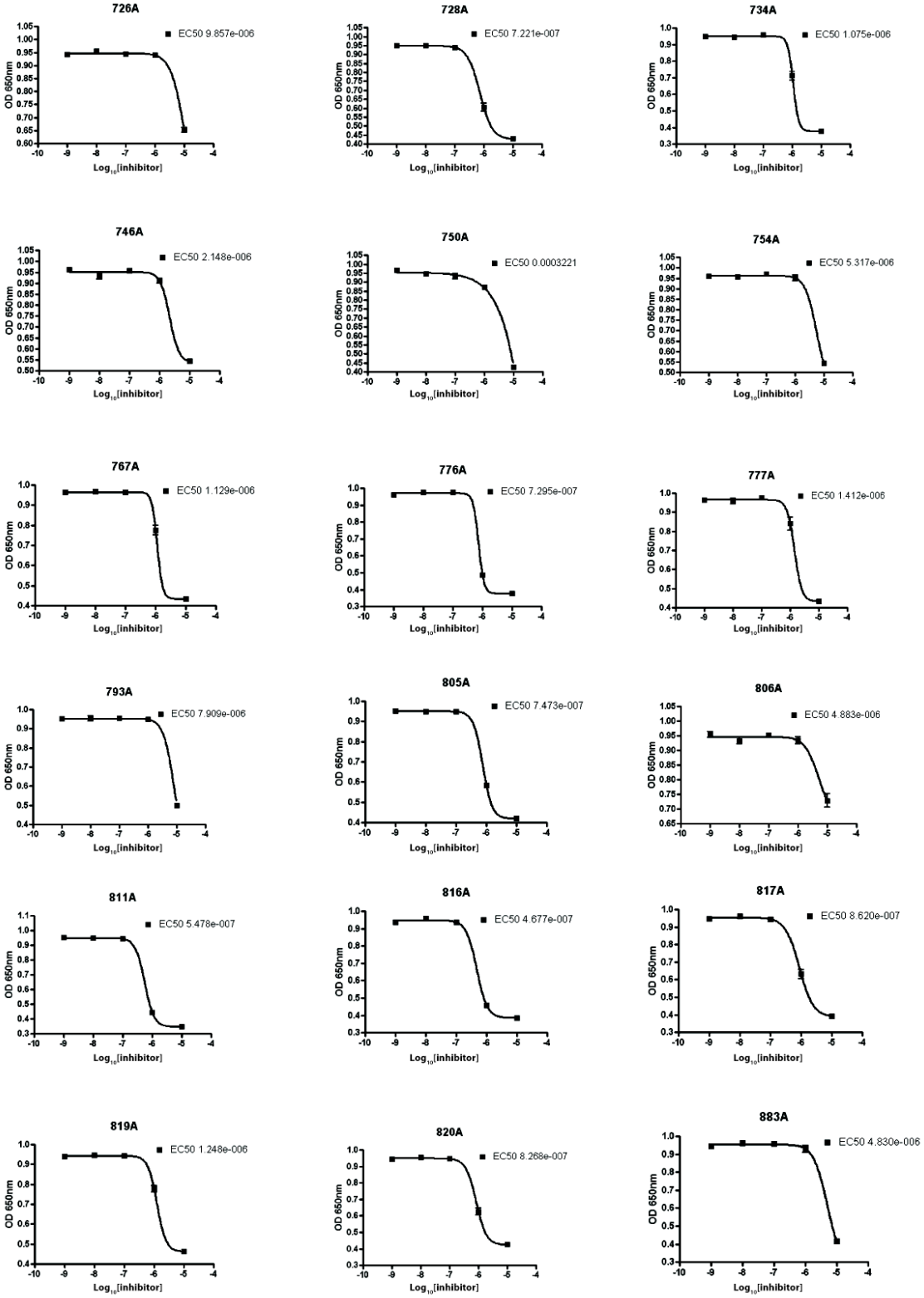


Figure 3.8. continued.

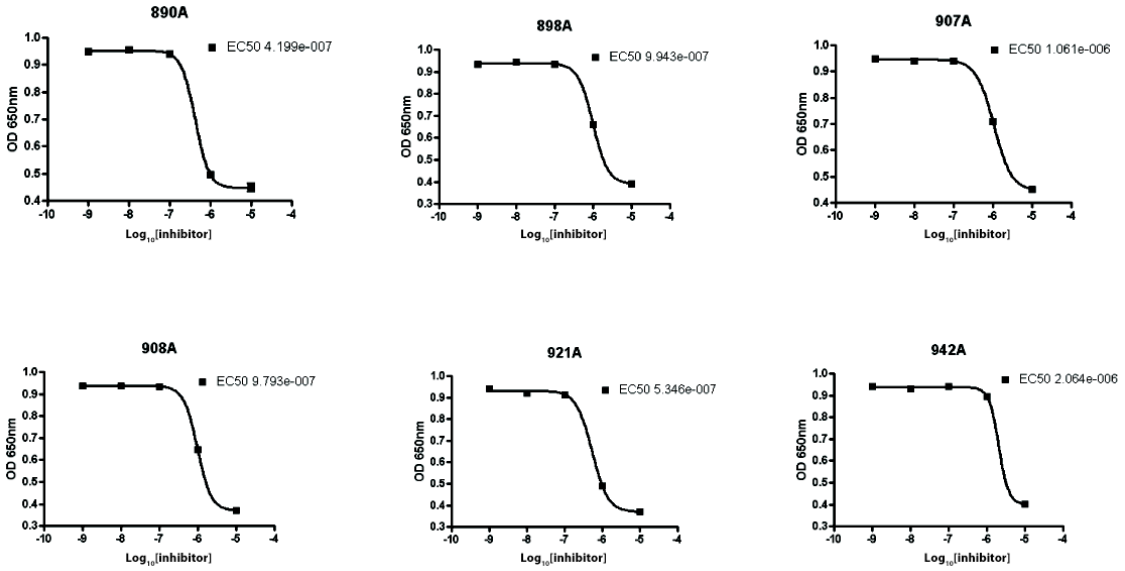


Figure 3.9. Structures of top 10 bisphosphonate hits.

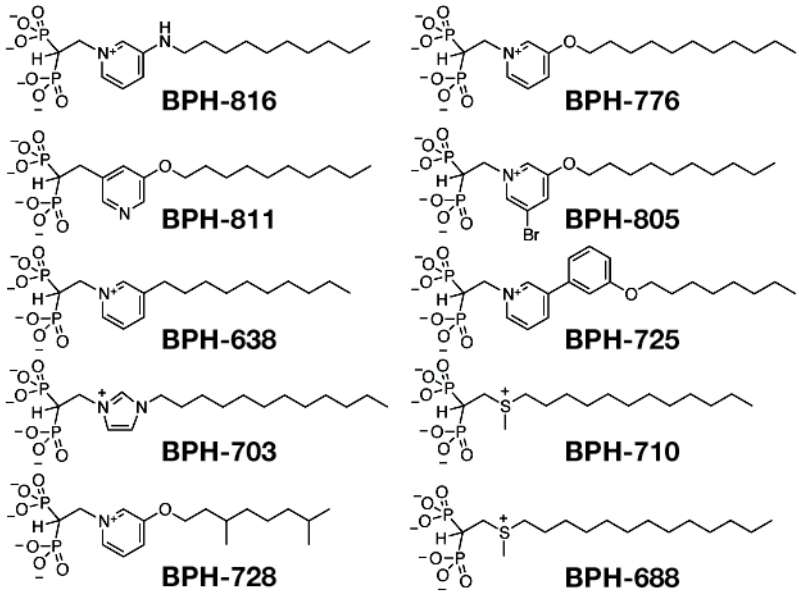


Figure 3.10. Rescue experiments on BPH-703 and BPH-811 by solanesol, FOH and GGOH.

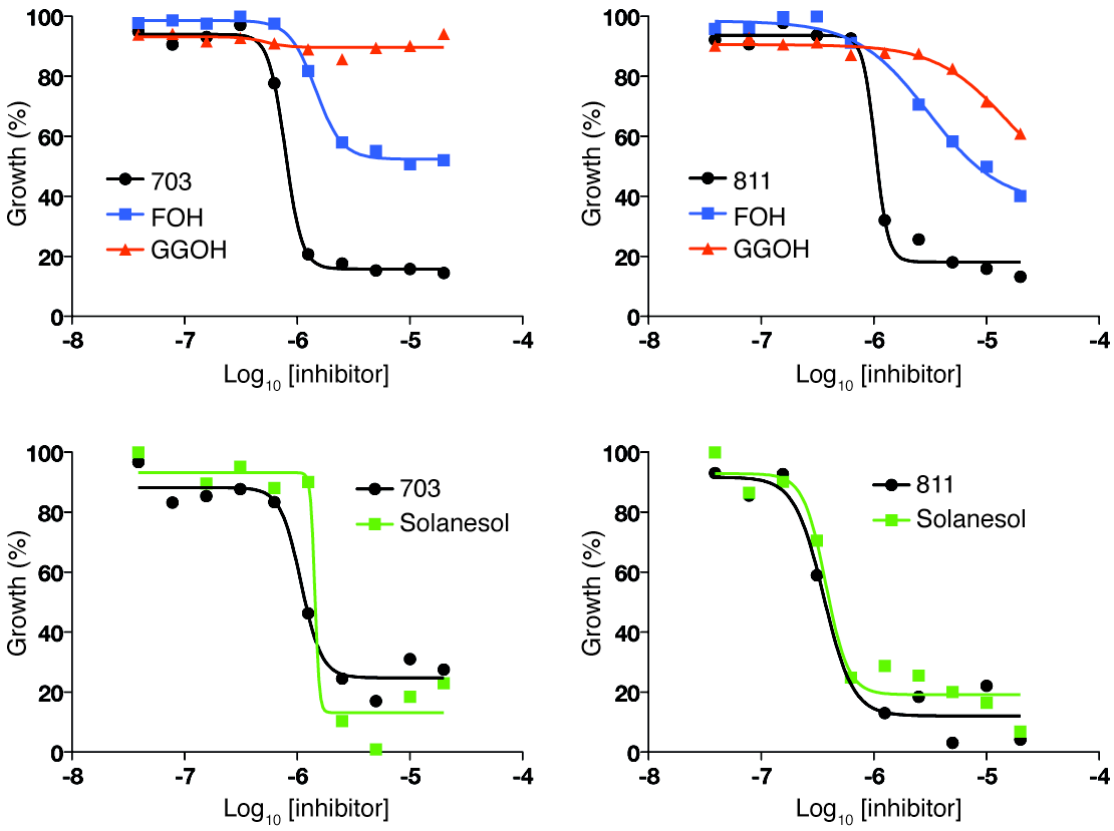


Figure 3.11. Surface representation of (**A**)zoledronate, (**B**) BPH-703, (**C**) BPH-811 and (**D**) GGPP.

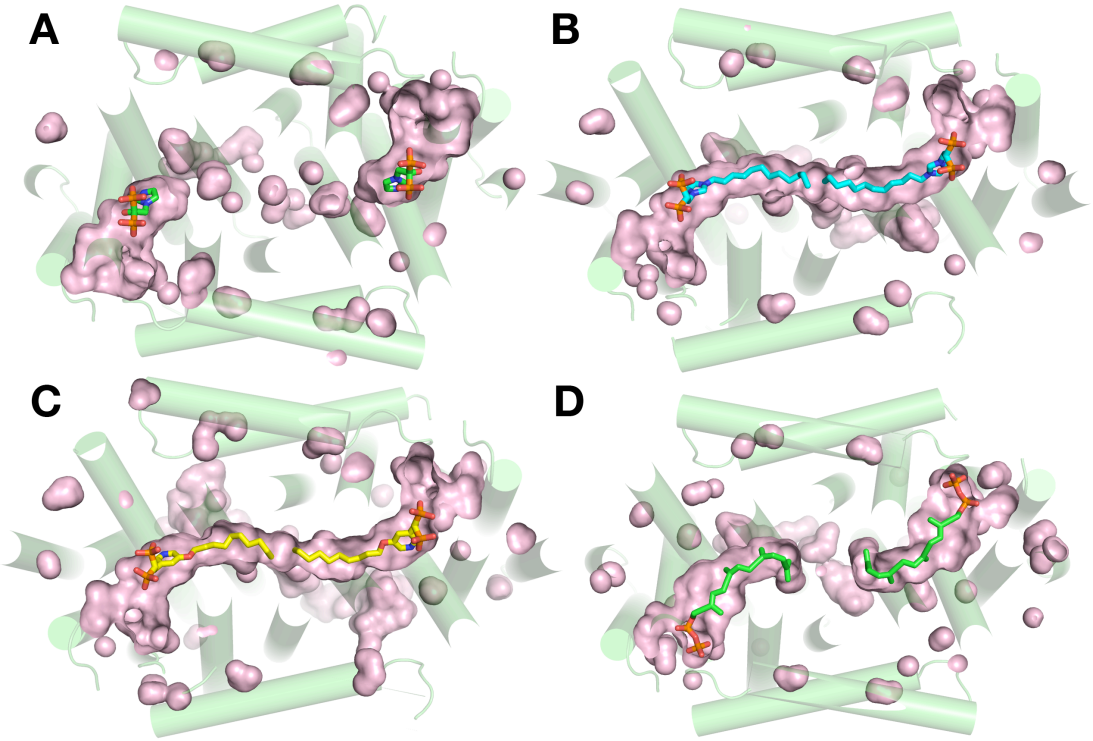
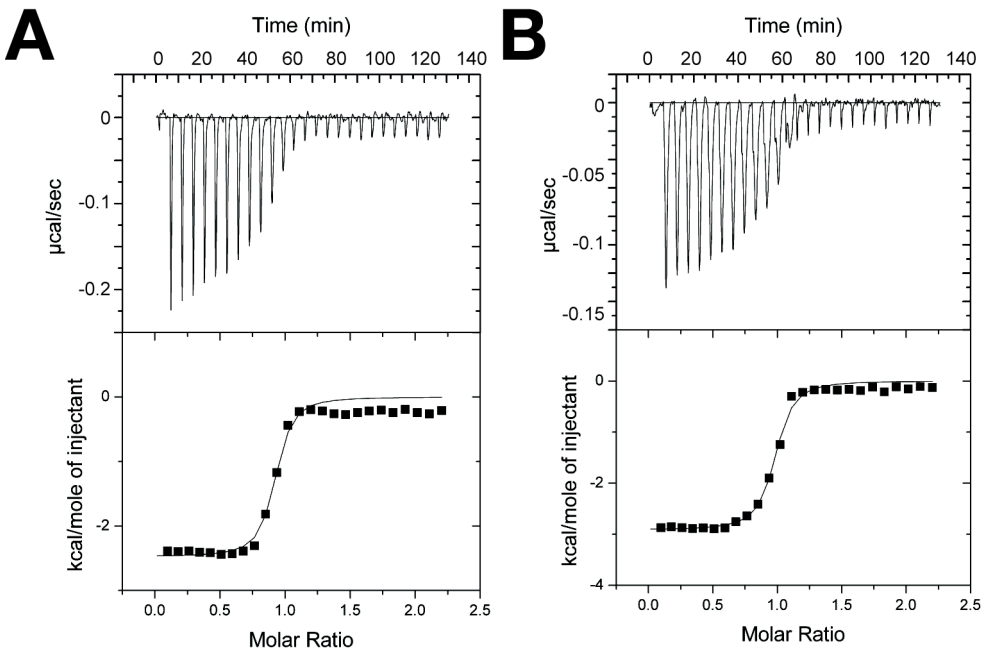


Figure 3.12. ITC results of (**A**) risedronate and (**B**) BPH-811.



3.7 References

1. WHO (2010) World Malaria Report 2010. (World health organization:, Geneva, Switzerland), p http://www.who.int/malaria/world_malaria_report_2010/en/index.html.
2. Dondorp AM, *et al.* (2010) Artemisinin resistance: current status and scenarios for containment. (Translated from eng) *Nat Rev Microbiol* 8(4):272-280 (in eng).
3. Ebetino FH, *et al.* (2011) The relationship between the chemistry and biological activity of the bisphosphonates. (Translated from Eng) *Bone* (in Eng).
4. Martin MB, *et al.* (2001) Bisphosphonates inhibit the growth of *Trypanosoma brucei, Trypanosoma cruzi, Leishmania donovani, Toxoplasma gondii,* and *Plasmodium falciparum*: a potential route to chemotherapy. *J Med Chem* 44(6):909-916.
5. Garzoni LR, *et al.* (2004) Selective in vitro effects of the farnesyl pyrophosphate synthase inhibitor risedronate on *Trypanosoma cruzi*. *Int J Antimicrob Agents* 23(3):273-285.
6. Montalvetti A, *et al.* (2003) Farnesyl pyrophosphate synthase is an essential enzyme in *Trypanosoma brucei*. *In vitro* RNA interference and *in vivo* inhibition studies. *J Biol Chem* 278(19):17075-17083.
7. Rodriguez N, *et al.* (2002) Radical cure of experimental cutaneous leishmaniasis by the bisphosphonate pamidronate. *J Infect Dis* 186(1):138-140.
8. Yardley V, *et al.* (2002) *In vivo* activities of farnesyl pyrophosphate synthase inhibitors against *Leishmania donovani* and *Toxoplasma gondii*. *Antimicrob Agents Chemother* 46(3):929-931.
9. Ling Y, *et al.* (2005) Bisphosphonate inhibitors of *Toxoplasma gondi* growth: *in vitro*, QSAR, and *in vivo* investigations. *J Med Chem* 48(9):3130-3140.

10. Moreno B*, et al.* (2001) (31)P NMR of apicomplexans and the effects of risedronate on *Cryptosporidium parvum* growth. *Biochem Biophys Res Commun* 284(3):632-637.

11. Artz JD*, et al.* (2008) Targeting a uniquely nonspecific prenyl synthase with bisphosphonates to combat cryptosporidiosis. *Chem Biol* 15(12):1296-1306.

12. Bruchhaus I, Jacobs T, Denart M, & Tannich E (1996) Pyrophosphate-dependent phosphofructokinase of *Entamoeba histolytica*: molecular cloning, recombinant expression and inhibition by pyrophosphate analogues. *Biochem J* 316 (Pt 1):57-63.

13. Ghosh S*, et al.* (2004) Effects of bisphosphonates on the growth of *Entamoeba histolytica* and Plasmodium species in vitro and in vivo. *J Med Chem* 47(1):175-187.

14. Singh AP*, et al.* (2010) Lipophilic bisphosphonates are potent inhibitors of Plasmodium liver-stage growth. (Translated from eng) *Antimicrob Agents Chemother* 54(7):2987-2993 (in eng).

15. Jordao FM*, et al.* (2011) *In Vitro* and *In Vivo* Antiplasmodial Activities of Risedronate and Its Interference with Protein Prenylation in *Plasmodium falciparum*. (Translated from eng) *Antimicrob Agents Chemother* 55(5):2026-2031 (in eng).

16. Artz JD*, et al.* (2011) Molecular characterization of a novel geranylgeranyl pyrophosphate synthase from Plasmodium parasites. (Translated from eng) *J Biol Chem* 286(5):3315-3322 (in eng).

17. Mukkamala D, No JH, Cass LM, Chang TK, & Oldfield E (2008) Bisphosphonate inhibition of a Plasmodium farnesyl diphosphate synthase and a general method for predicting cell-based activity from enzyme data. *J Med Chem* 51(24):7827-7833.

18. Mukherjee S, Song Y, & Oldfield E (2008) NMR investigations of the static and dynamic structures of bisphosphonates on human bone: a molecular model. *J Am Chem Soc* 130(4):1264-1273.

19. Mukherjee S, Huang C, Guerra F, Wang K, & Oldfield E (2009) Thermodynamics of bisphosphonates binding to human bone: a two-site model. *J Am Chem Soc* 131(24):8374-8375.

20. Zhang Y*, et al.* (2009) Lipophilic bisphosphonates as dual farnesyl/geranylgeranyl diphosphate synthase inhibitors: an X-ray and NMR investigation. *J Am Chem Soc* 131(14):5153-5162.

21. Zhang Y*, et al.* (Lipophilic pyridinium bisphosphonates: potent gammadelta T cell stimulators. *Angew Chem Int Ed Engl* 49(6):1136-1138.

22. Liu CI*, et al.* (2008) A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* 319(5868):1391-1394.

23. Song Y*, et al.* (2009) Phosphonosulfonates are potent, selective inhibitors of dehydrosqualene synthase and staphyloxanthin biosynthesis in *Staphylococcus aureus*. *J Med Chem* 52(4):976-988.

24. Song Y*, et al.* (2009) Inhibition of staphyloxanthin virulence factor biosynthesis in *Staphylococcus aureus*: *in vitro*, *in vivo*, and crystallographic results. *J Med Chem* 52(13):3869-3880.

25. Jomaa H*, et al.* (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285(5433):1573-1576.

26. Chakrabarti D*, et al.* (1998) Protein prenyl transferase activities of *Plasmodium falciparum*. *Mol Biochem Parasitol* 94(2):175-184.

27. Tonhosolo R*, et al.* (2009) Carotenoid biosynthesis in intraerythrocytic stages of *Plasmodium falciparum*. *J Biol Chem* 284(15):9974-9985.

28. Tonhosolo R*, et al.* (Intraerythrocytic stages of *Plasmodium falciparum* biosynthesize menaquinone. *FEBS Lett* 584(23):4761-4768.

29. D'Alexandri FL, Gozzo FC, Eberlin MN, & Katzin AM (2006) Electrospray ionization mass spectrometry analysis of polyisoprenoid alcohols via Li+ cationization. *Anal Biochem* 355(2):189-200.

30. C KMC*, et al.* (2008) Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates: a crystallographic and computational investigation. *J Med Chem* 51(18):5594-5607.

31. Lin FY*, et al.* (2010) Mechanism of action and inhibition of dehydrosqualene synthase. (Translated from eng) *Proc Natl Acad Sci U S A* 107(50):21337-21342 (in eng).

32. Cammerer SB*, et al.* (2007) Quinuclidine derivatives as potential antiparasitics. *Antimicrob Agents Chemother* 51(11):4049-4061.

33. Zhang JH, Chung TD, & Oldenburg KR (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. (Translated from Eng) *J Biomol Screen* 4(2):67-73 (in Eng).

34. Martin MB*, et al.* (2002) Activity of bisphosphonates against *Trypanosoma brucei* rhodesiense. *J Med Chem* 45(14):2904-2914.

35. Kavanagh KL, Dunford JE, Bunkoczi G, Russell RG, & Oppermann U (2006) The crystal structure of human geranylgeranyl pyrophosphate synthase reveals a novel hexameric arrangement and inhibitory product binding. *J Biol Chem* 281(31):22004-22012.

36. Guo RT, *et al.* (2007) Bisphosphonates target multiple sites in both cis- and trans-prenyltransferases. *Proc Natl Acad Sci U S A* 104(24):10022-10027.

37. Rondeau JM, *et al.* (2006) Structural basis for the exceptional *in vivo* efficacy of bisphosphonate drugs. *ChemMedChem* 1(2):267-273.

38. Kavanagh KL, *et al.* (2006) The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *Proc Natl Acad Sci U S A* 103(20):7829-7834.

39. Yin F, Cao R, Goddard A, Zhang Y, & Oldfield E (2006) Enthalpy versus entropy-driven binding of bisphosphonates to farnesyl diphosphate synthase. *J Am Chem Soc* 128(11):3524-3525.

40. Makler MT, *et al.* (1993) Parasite lactate dehydrogenase as an assay for *Plasmodium falciparum* drug sensitivity. (Translated from eng) *Am J Trop Med Hyg* 48(6):739-741 (in eng).

41. Smilkstein M, Sriwilaijaroen N, Kelly JX, Wilairat P, & Riscoe M (2004) Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. (Translated from eng) *Antimicrob Agents Chemother* 48(5):1803-1806 (in eng).

42. Zhang Y, *et al.* (2009) Lipophilic bisphosphonates as dual farnesyl/geranylgeranyl diphosphate synthase inhibitors: an X-ray and NMR investigation. (Translated from eng) *J Am Chem Soc* 131(14):5153-5162 (in eng).

43. Zhang Y, *et al.* (2007) Activity of sulfonium bisphosphonates on tumor cell lines. (Translated from eng) *J Med Chem* 50(24):6067-6079 (in eng).

44. Sanders JM, *et al.* (2005) Pyridinium-1-yl bisphosphonates are potent inhibitors of farnesyl diphosphate synthase and bone resorption. (Translated from eng) *J Med Chem* 48(8):2957-2963 (in eng).

45. Mukkamala D, No JH, Cass LM, Chang TK, & Oldfield E (2008) Bisphosphonate inhibition of a Plasmodium farnesyl diphosphate synthase and a general method for predicting cell-based activity from enzyme data. (Translated from eng) *J Med Chem* 51(24):7827-7833 (in eng).

46. Schrodinger (2011) Maestro version 9.2 New York, NY).

47. Schrodinger (2011) LigPrep, version 2.5 New York, NY).

48. Schrodinger (2011) Glide, version 5.7 New York, NY).

49. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. (Translated from English) *Method Enzymol* 276:307-326 (in English).

50. McCoy AJ*, et al.* (2007) Phaser crystallographic software. (Translated from Eng) *J Appl Crystallogr* 40(Pt 4):658-674 (in Eng).

51. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. (Translated from English) *Acta Crystallogr D* 60:2126-2132 (in English).

52. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. (Translated from English) *Acta Crystallogr D* 53:240-255 (in English).

53. Potterton E, Briggs P, Turkenburg M, & Dodson E (2003) A graphical user interface to the CCP4 program suite. (Translated from English) *Acta Crystallogr D* 59:1131-1137 (in English).

54. Brunger AT*, et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. (Translated from English) *Acta Crystallogr D* 54:905-921 (in English).

55. Schuttelkopf AW & van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. (Translated from English) *Acta Crystallogr D* 60:1355-1363 (in English).

56. DeLano WL (2008) The PyMOL Molecular Graphics Systems. (DeLano Scientific LLC, Palo Alto, CA, USA.).

Chapter 4

Lipophilic Bisphosphonates Are Potent Inhibitors of Plasmodium Liver-Stage Growth

4.1 Notes and Acknowledgements

This work was supported by the Dana Foundation (to Victor Nussenzweig) and the United States Public Health Service (NIH grant GM65307, to Eric Oldfield) and the Barbara and Sanford Orkin Chair in Tropical and Emerging Diseases (to Roberto Docampo). We thank Dr. Raymond Hui for providing the *P. vivax* GGPPS expression system; K. Kavanagh and U. Oppermann for providing the human GPPS expression system and H. Sagami for providing the human GGPPS expression system. We also thank Patric J. Casey for providing compound DU40.

4.2 Introduction

The malaria parasite *Plasmodium* is transmitted via the bite of an infected female *Anopheles* mosquito that injects sporozoites into the skin. The parasites enter the bloodstream and invade hepatocytes, where the sporozoites transform into liver stages, usually called exo-erythrocytic forms (EEFs). These then undergo multiple rounds of nuclear division and produce thousands of merozoites that enter the blood circulation and infect erythrocytes, leading to malaria symptoms. EEFs are thus targets for prophylactic drugs and vaccines. At present, however, there are a very limited number of drugs that target *Plasmodium* liver stages. Primaquine is one, but its use is limited by its associated toxicity and increased risk of hemolysis when administered to persons with glucose-6-phosphate dehydrogenase (G6PD) deficiency (2). This latter problem also affects

two related compounds under development: bulaquine and tafenoquine (19, 20). Although there has been considerable progress in the treatment of bloodstream infections using artemisinin-based combination therapies, recent reports have indicated the possible development of resistance to artemisinin among Asian isolates (6). This situation is potentially serious considering how widespread is the resistance to the other anti-malarials, chloroquine, or the pyrimethamine-sulfadoxine combination (14). Under these circumstances, there is, therefore, a need to identify new lead compounds active against malaria. In particular, there is currently great interest (10) in developing drugs that inhibit the liver stage forms that precede blood infections, thereby preventing symptomatology, and blocking transmission (21).

During the early stages of EEF development, the *Plasmodium* plasma membrane is covered with the circumsporozoite protein (CSP). In infected hepatocytes, parasites are contained inside a parasitophorous membrane, but CSP is still exported into the hepatocyte cytoplasm and nucleus (17), generating profound changes in the transcriptional program of the host cell. In previous work we noticed that the gene that was most highly upregulated on CSP expression in infected HeLa cells (17) or HepG2 cells (M. Zhang, and V. Nussenzweig, unpublished data) was Rap1A. Since Rap proteins are geranylgeranylated at their C-terminus, and since there are such large changes in Rap1A gene expression on CSP expression in HeLa cells, it seemed possible that blocking protein geranylgeranylation would have a significant effect on liver stage parasite proliferation.

To inhibit prenylation we investigated the effects of several different bisphosphonates, a class of drugs known to inhibit protein prenylation. Bisphosphonates are potent, low nM inhibitors of the enzyme farnesyl diphosphate synthase (FPPS), which catalyzes the condensation of the isoprenoids dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP),

produced in the mevalonate pathway (Figure 4.1) to form geranyl diphosphate (GPP). This then condenses with a second IPP molecule to form farnesyl diphosphate (FPP) and thence, geranylgeranyl diphosphate (GGPP) (Figure 4.1). As a result of this FPPS inhibition, bisphosphonates block protein prenylation, as well as sterol, ubiquinone, dolichol, and heme *a* biosynthesis (15). Bisphosphonates also stimulate human γδ T cells (containing the Vγ2Vδ2 T cell receptor) (8), and may therefore have anti-parasitic activity. Bisphosphonate inhibition of FPPS also results in the accumulation of the substrate IPP, which becomes conjugated to AMP to form a toxic ATP analogue, O-isopentenyl-ATP, ApppI, Figure 4.1, which can inhibit the mitochondrial adenine nucleotide translocase and induce apoptosis (11).

We reasoned that novel, lipophilic bisphosphonates (23) might have enhanced effects on the survival of malaria parasites over the more polar bisphosphonates reported previously (7, 9) and we focus here on a comparison of the activity of three lipophilic bisphosphonates (currently being developed as anti-cancer drug leads) with the activity of the second and third generation bisphosphonates pamidronate and risedronate, in both liver stage (EEF) and blood stage *Plasmodium* infection models.

4.3 Results and Discussion

Growth inhibition of *P. berghei* EEFs *in vitro*.

We first tested the effects of two bisphosphonate drugs, pamidronate (Aredia) and risedronate (Actonel) (Figure 4.2), together with the effects of pravastatin, an isoprenoid biosynthesis inhibitor that targets 3-hydroxymethyl-glutaryl-coenzyme A (HMG-CoA) reductase (Figure 4.1, 4.2), on *P. berghei* EEF growth in HepG2 cells. Pamidronate and pravastatin were inactive at the concentrations tested (IC_{50} >200 μM), Table 4.1. The risedronate IC_{50} was ~16 μM (Figure 4.3B

and Table 4.1) while the lipophilic bisphosphonates (BPH-715; BPH-942 and BPH-943; Table 4.1 and Figure 4.3A) had IC_{50} values in the range ~8-10 μM. We also tested the toxicity of the active bisphosphonates on HepG2 cells and their TC_{50}s were in the range of 1-13 mM (Table 4.1). All of the active bisphosphonates had therapeutic index values >100, comparable to that of primaquine (5). In addition, we tested the geranylgeranyl transferase I inhibitor GGTI-298 (Figure 4.2) finding an IC_{50} ~30 μM (Table 4.1), but this compound was later found to have no activity *in vivo* (Table 4.2).

Bisphosphonates inhibit the growth of *P. berghei* EEFs at an early stage.

We next treated sporozoite-infected hepatocytes with one of the lipophilic bisphosphonates (BPH-715), chosen since it has already been tested for safety in mice (23) and examined the cultures 50 hours later, using light microscopy, Figure 4.4A, B. When BPH-715 was present at 30 μM, there was a >9x reduction in the EEF area (Figure 4.4C). There was no difference in the total numbers of EEFs in the control and treated groups. BPH-715 thus inhibits EEF growth at an early trophozoite stage.

Bisphosphonates also inhibit *P. berghei* EEF growth *in vivo*.

To evaluate to what extent the lipophilic bisphosphonates had activity *in vivo*, we injected them i.p. once daily into mice, starting two days before and ending two days after, infection with sporozoites (a total of 5 days). When challenged with 3,000 infectious sporozoites, the bisphosphonates showed a wide range of effects, ranging from no activity to complete protection (Table 4.2, Tables 4.3 and 4.4). The most active compound (BPH-715) protected mice completely at a dose of 1.5 mg/kg body weight (Table 4.2 and Table 4.4). The next most active compound, BPH-943, showed a five-day delay in the pre-patent period at a dose of 2 mg/kg body weight, which also represents a very large reduction in parasite burden. Risedronate was less

active than was BPH-715, although it did show a 4-day reduction in pre-patent period at 20 mg/kg (Table 4.2). The lipophilic bisphosphonates were, therefore, far more potent than the commercial bisphosphonate, risedronate. We also tested GGTI-298 (Figure 4.2), a commercially available inhibitor of geranylgeranyl transferase (GGTI) *in vivo* (up to a dose of 5 mg/kg), but GGTI-298 did not show any activity against the liver stage parasites *in vivo*, although it did show very modest (~30 μM) activity *in vitro*. Another GGTI inhibitor (DU40, Figure 4.2) also failed to show any *in vivo* activity against the EEFs. So, only the lipophilic bisphosphonates have good *in vivo* activity.

Bisphosphonates inhibit *P. falciparum* blood stage parasites *in vitro.*

We next tested risedronate and the three lipophilic bisphosphonates for activity against *P. falciparum*, *in vitro*. We found that with the chloroquine sensitive strain 3D7, the lipophilic bisphosphonates (BPH-715, BPH-942 and BPH-943) had IC_{50} values in the sub-micromolar range (Table 4.5), while risedronate had an IC_{50} of ~1.2 μM (Table 4.5). BPH-942 and BPH-943 showed 3-4-fold less activity against the chloroquine-resistant strain (W2mef) as compared to the chloroquine sensitive strain (3D7), but BPH-715 had good activity, against both strains (Table 4.5).

Bisphosphonates inhibit *P. berghei* blood stage parasites *in vivo.*

Two bisphosphonates (risedronate and the lipophilic bisphosphonate, BPH-715) were then tested *in vivo* for activity against a blood stage parasite challenge. Risedronate was given at 20 mg/kg for 5 days with the mice being challenged with 5×10^6 blood stage parasites on day 0. Under these conditions, there was a one day delay in the pre-patent period (Table 4.6). BPH-715, on the other hand, was more effective. A similar treatment regime with BPH-715 at 1.5 mg/kg gave

a 4 day delay in the pre-patent period, indicating a larger reduction in blood stage parasite growth.

Activity of bisphosphonates against *Plasmodium* GGPPS and against human FPPS and GGPPS.

In order to better understand the activities of the different bisphosphonates *in vitro* and *in vivo*, we determined their IC_{50} values against three different drug targets: a *Plasmodium vivax* GGPPS (using GPP, FPP as substrates); human FPPS (HsFPPS) and human GGPPS (HsGGPPS) using the methods described previously (13, 16, 22, 23) . Results are shown in Table 4.7. As can be seen in Table 4.7, all compounds inhibit the PvGGPPS as well as HsFPPS. However, only the lipophilic bisphosphonates inhibit HsGGPPS, with risedronate being inactive ($IC_{50} \gtrsim 300$ μM, Table 4.7). These results are consistent with direct activity against the *Plasmodium* enzyme, which can be expected to contribute to their activity in the cell-based assays. What is puzzling, however, is that the pattern of activity with the lipophilic bisphosphonates (BPH-715, BPH-942 and BPH-943) against either the PvGGPP, or HsFPPS and HsGGPPS, does not track the activity seen *in vivo*. *In vitro*, in both the EEF and blood stage assays, all three compounds have broadly similar activity (~8-10 μM, *in vitro*, EEF, Table 4.1; ~0.5-2.2 μM, *in vitro*, blood stage assay) and this is consistent with their similar range of activity in the enzyme inhibition assays, Table 4.7. However, in the *in vivo* EEF experiments (Table 4.2), BPH-715 is clearly far more effective than is BPH-943 (at essentially the same dosing: 1.5 mg/kg vs. 2.0 mg/kg). A likely explanation for this is that BPH-715 has much better uptake into the liver since it is more lipophilic than is BPH-943. Table 4.7 shows the SlogP (the logarithm of the oil/water partition coefficient) values computed in MOE (1) for all four bisphosphonates. From these results we see that BPH-715 is clearly more lipophilic than is BPH-943, with an SlogP value of -1.2, to be compared with -3.4

for BPH-943. The -5.5 SlogP for risedronate represents an extremely hydrophilic species, which together with its lack of activity against HsGGPPS (Table 4.7) likely explains its lower activity than the lipophilic bisphosphonates, *in vivo*.

4.4 Conclusions

The results we have described above show that lipophilic bisphosphonates have potent activity against liver stage *P. berghei in vivo*, with the most active compound completely protecting mice against sporozoite challenge (at a dose of 1.5 mg/kg body weight for 5 days) over the 28 day observation period, with no adverse effects. The enhanced activity of this compound over that seen with the other bisphosphonates tested arises, we propose, from enhanced uptake into the liver. The bisphosphonate risedronate has poorer efficacy than does BPH-715 *in vivo*, but still provides a modest delay (4 days) in the pre-patent period. Interestingly, while BPH-715 is more potent in inhibiting liver stage rather blood stages *in vivo*, it generates a much smaller effect on blood stage parasites *in vivo*, even though it is quite potent *in vitro*. One possible explanation is that BPH-715 uptake into the liver may simply be much more effective because of the much higher lipophilicity of this species (SlogP value 2-4 units greater than with any of the other bisphosphonates tested). At present, we cannot quantify the relative importance of host *versus* pathogen targeting in the liver stages, where PbGGPPS, HsFPPS and/or HsGGPPS could all be inhibited, to varying degrees. Indeed, in recent work (13) we reported that, in TEN simpler systems, the correlation between enzyme inhibition ($pIC_{50} = -\log_{10} IC_{50}$) and cell activity pIC_{50} values for a broad range of assays, was only $R^2 = 0.30$ (13). This pointed to a key role of transport or cell permeability and when this was taken into account, the correlation improved to $R^2 = 0.70$ (13). However, applying the mathematical model used in that approach to liver stage

Plasmodium spp. (*in vitro* or *in vivo*) is a challenge since there are, in principle, at least 3 targets (PbGGPPS, HsFPPS and HsGGPPS) and at least THREE "permeability" barriers (the host cell, the *Plasmodium* plasma membrane, and the parasitophorous vacuole membrane). In blood stage Plas*modium*, this may be more tractable since the HsFPPS/HsGGPPS are not targets, but quite large data sets (>20 compounds) are needed for this sort of analysis (13). In addition, we cannot rule out other possible targets, such as octaprenyl diphosphate synthase, and indeed, in other work we have shown that bisphosphonates can inhibit multiple prenyl synthase targets- something that may contribute to their potent activity.

In any case, the results we have obtained here are clearly of considerable interest since we find complete protection against *P. berghei* liver stage infection in mice using the lipophilic bisphosphonate, BPH-715. Interestingly, this compound has also been found to be the most potent inhibitor of tumor cell growth, both *in vitro* and *in vivo* (23), due, we proposed, to its potent activity against both human FPPS and GGPPS, and its lipophilic nature (as compared with other bisphosphonates). This suggests the possibility that BPH-715 does, in fact, block protein geranylgeranylation and notably, in three human cell lines, we found that cell growth inhibition could be "rescued" by addition of geranylgeraniol – although of course this rescue-effect would not have probative value in *Plasmodium* liver stage experiments since the effects on both PbGGPPS and HsGGPPS inhibition would both be "rescued" (by providing the exogenous isoprenoid source).

In the blood stage *P. falciparum* experiments, the effects of BPH-715 and the more hydrophilic species, risedronate, are quite similar to each other, with the IC_{50} ratio (risedronate/BPH-715) being only about a factor of 2-3. Since there is no isoprenoid biosynthesis in the red blood cell, HsFPPS/HsGGPPS inhibition is not involved, so solely the

Plasmodium GGPPS is being targeted. Notably, this factor of ~2.5 is the same as that seen with the *Plasmodium* EEFs (Table 4.1) in which risedronate has an IC_{50} of 16 μM for EEF inhibition, while that for BPH-715 is IC_{50} = 10 μM (Table 4.1), suggesting the possibility that the *Plasmodium* GGPPS is a primary target in both blood and liver stage forms, *in vitro*. The improved activity of the more lipophilic bisphosphonates is in agreement with results we reported previously (on much less active compounds): that the most lipophilic bisphosphonates have the best activity, against *Plasmodium* blood stages parasites (7).

In summary then: lipophilic bisphosphonates are potential leads against the liver stages of malaria parasites, including the hypnozoites of *P. vivax*, and could have a role in malaria chemoprophylaxis.

4.5 Materials and Methods

Animals.

The methods for mice use and care were approved by the New York University School of Medicine Institutional Animal Care and Use Committee and are in accordance with PHS policy on Care and Use of Laboratory Animals under the guidance of the Office of Laboratory Animal Welfare (OLAW) within the NIH.

Reagents.

The compounds used were: pravastatin and pamidronate from Sigma; GGTI-298 from Calbiochem; risedronate, synthesized as described in (9); and BPH-715, BPH-942, and BPH-943, lipophilic bisphosphonates synthesized as described (4) .

***In vitro* growth inhibition assay for liver stage parasites.**

Drug solutions were prepared in 1x sterile PBS. The diluted solutions were added to 24-well

culture plates containing human HepG2 cells seeded a day prior to the experiment, and 0.5 ml complete DMEM medium containing 10% FBS, together with an antibiotic-antimycotic. Infection was initiated by adding 10,000 *P. berghei* sporozoites (NK65 line). Infected cultures were then allowed to grow at 37°C in a 5 % CO_2 atmosphere, for 48 hours. Culture medium was changed 24 hours after infection, and fresh compounds were added at the same concentration, to maintain inhibitor pressure throughout the growth period. At the end of the 48 hour incubation period, total RNA was extracted using Trizol™ reagent. Reverse transcription from 4 μg RNA was performed to obtain cDNA. In a real-time PCR mix of 50 μL, cDNA equivalent of 0.5 μg RNA was used. The real-time PCR mix also contained *P. berghei* 18S rRNA specific primers and SYBR green dye (3). Real time PCR was performed in a BioRad Icycler and the copy numbers were calculated, using the known amount of plasmid standard having the amplification target sequence (3). Parasite growth inhibition was calculated by dividing the 18S rRNA copy number of the experimental group by that of the untreated control group. Non-linear regression analysis was performed using Sigma-plot (version-10), to determine IC_{50} values.

***In vivo* growth inhibition assay for liver stage parasites (EEFs).**

Compounds were injected into mice i.p. for a variable number of days, depending on the experiment. Treatments started on day -2, -1 or 0, where day 0 corresponds to sporozoite infection by the intravenous route. Treatments were continued on day +1 and +2 post infection. EEF development was evaluated by the success of the ensuing blood stage infection. Blood smears were prepared on day three, post-infection. Percent parasitemia was monitored until it reached 3-5%, and animals were then sacrificed. Complete protection is defined as the absence of parasites in blood smears up to two weeks post challenge with sporozoites.

Blood stage parasite growth inhibition assay in infected mice.

Mice (five per group) were injected i.p. with $5x10^6$ blood stage *P. berghei* parasites/mouse, to initiate infection. Treatments started on day 0 and were continued on days +1 and +2, post infection. Blood smears were prepared on day 2 and onwards, post parasite-challenge. Parasitemia was monitored microscopically until it reached 3-5%, then the animals were sacrificed.

***In vitro* growth inhibition assay for *P. falciparum* blood stage parasites.**

Bisphosphonates were evaluated for their inhibitory activity against blood stages of *P. falciparum* strains 3D7 (chloroquine sensitive) and W2mef (chloroquine resistant). Each compound was tested in triplicate in 96-well culture plates in a volume of 100 µL. Drug concentrations of 1, 10, 100, 1,000, and 10,000 nM were tested. Chloroquine at concentrations of 10 and 100 nM was used as a positive control. *P. falciparum* were cultured in O-positive washed erythrocytes in RPMI 1640 medium using the candle-jar technique (18). Parasites were added to reach an appropriate parasitemia (0.2-0.3%). Flasks were kept in a dessicator, and a low oxygen environment was produced by flushing with mix gas (90% N_2, 5% CO_2 and 5% O_2). Cells were incubated at 37°C and percent parasitemia determined microscopically.

Toxicity assays in hepatocytes.

HepG-2 (human liver hepatoma) cells were seeded at 70% confluency (with 0.5 mL medium) in 24-well tissue culture plates a day prior to treatment with inhibitors. Various concentrations of inhibitors were added to the HepG2 cells, which were then incubated at 37°C under 5 % CO_2 for 22 hours. After incubation, cell titer one-reagent^{TM} (Promega) was added to the wells, followed by 2 additional hours of incubation (12). The plates were briefly spun and the clear media transferred to 96-well plates. The OD_{495} was then determined by using a microplate reader

(Molecular Devices), and the 50% cell growth inhibition (IC_{50}) value determined by nonlinear regression (percent viable cells at various concentrations of the individual BPs).

Immunofluorescence staining and size measurements.

HepG 2 cells were seeded (50% confluent) into 24 well tissue culture plates containing glass cover slips, a day prior to bisphosphonate treatment. BPH-715 was added one hour prior to the addition of *P. berghei* sporozoites. Cells were then incubated at 37 ·C and 5% CO_2 for 48 hours. Cover slips were removed after 48 hours incubation and cells were fixed with 4% formaldehyde, followed by permeabilization with saponin [0.2 % in phosphate buffered saline (PBS)]. Permeabilized cells were blocked with 3% bovine serum albumin (BSA) in PBS for 2 hours, followed by 1-hour incubation with primary antibody 2E6 (monoclonal, against parasite HSP70). After the primary antibody incubation, three washes with PBST (PBS with 0.05% Tween-20) were given in order to diminish nonspecific binding. Cells were further incubated with secondary antibody (rabbit anti-mouse IgG coupled to Alexa 594), for 1 hour, followed by three more washes with PBST. A final wash was given with PBS and cover slips were mounted using anti-fade (Bio-Rad) followed by sealing with nail polish. Images of liver stage parasites were captured in a CCD camera attached to a fluorescence microscope (Nikon E-800). The Image Pro-Plus program (Media Cybernetics, USA) was used to measure the size of the liver-stage parasites.

Enzyme inhibition assays.

We used exactly the same methods as described previously (13, 16, 23) for PvGGPPS, HsFPPS and HsGGPPS expression, purification, and inhibition.

4.6 Tables and Figures

Table 4.1. *In vitro* activity of compounds against *Plasmodium* EEFs.

Compound	IC_{50} for EEFs (μM)	TC_{50} for HepG2 cells (mM)	Therapeutic Index (TI)
Pamidronate	>200	ND[a]	NA[b]
Risedronate	16	13	813
Pravastatin	>200	ND	NA
BPH-715	10	2	200
BPH-942	8.8	1	114
BPH-943	8	1	125
GGTI-298	~30	ND[a]	NA[b]

[a] ND: Not determined
[b] NA: Not applicable.

Table 4.2. *In vivo* activity of bisphosphonates against *Plasmodium* EEFs.

Compounds	Mouse Dose (mg/kg) for 5 days (-2,-1,0, +1+2)	Pre-patent day [a]	Delay in pre-patent period (PPP) in days
Untreated	None	4	0
Risedronate	20	8	4
BPH-715	1.5	> 28[b]	> 28[b]
BPH-942	0.8	7	3
BPH-943	2	9	5
GGTI-298	5	4	0

[a] Day 0 corresponds to challenge with 3000 *P. berghei* sporozites.
[b] > 28 = No blood stage parasites through the 28-day observation period.

Table 4.3. *In vivo* drug testing (Risedronate): *P. berghei* sporozoite challenge.

Treatment (dose & days)	**Challenge (no. of sporozoites)**	**Difference In PPP**	**Mouse & age**	**Detection method & comments**
5mpk x 2 d (0,+1)	10,000	0.5 day	Balb/c 12wk	Smear 5 fold less parasite
17.5mpk x 2d (0,+1)	10,000	1 day	Balb/c 12wk	Smear 10 fold less parasite
20mpk x4d (0,+1+2+3)	3,000	4 days	C57BL/6 8 WK	Smear
20mpk x5d -1,0,+1,+2,+3	3,000	6 days	C57BL/6 8 WK	Smear (Delayed Toxicity)
20mpk x4d -2-1,0,+1,	3,000	2.5 days	C57BL/6 12 WK	Smear
20mpk x3d -1,0,+1,	3,000	0.5 day	C57BL/6 12 WK	Smear
10mpk x4d -2-1,0,+1,	3,000	No difference	C57BL/6 12 WK	Smear
10mpk x3d -1,0,+1,	3,000	No difference	C57BL/6 12 WK	Smear
20mpk x5d -2,-1,0,+1,+2,	10,000	3 days	C57BL/6 14 WK	Smear

Table 4.4. *In vivo* drug testing *P. berghei* sporozoite challenge.

Drug used	Treatment Dose & days	Challenge	Difference in Pre-patent period	Mouse & age	Detection method
BPH-715	0.5mpk x4d -2-1,0,+1,	3,000	3 day	C57BL/6 12 WK	Smear
GGTI-298	5mpk x5d -2,-1,0,+1,+2,	10,000	No difference	C57BL/6 14 WK	Smear
BPH-715	1.5mpk x5d -2,-1,0,+1,+2,	10,000	> 28 days No Blood Stages	C57BL/6 14 WK	Smear
BPH-715	5mpk x5d -2,-1,0,+1,+2,	10,000	> 28 days No Blood Stages	C57BL/6 14 WK	Smear

Table 4.5. IC_{50} values for the selected bisphosphonates against blood stage *Plasmodium falciparum* chloroquine-sensitive and-resistant strains *in vitro*.

Inhibitor	3D7, IC_{50} (nM) [Average ± SD]	W2mef, IC_{50} (nM) [Average ± SD]
Risedronate	1200 ± 90	1400 ± 100
BPH-715	670 ± 40	590 ± 15
BPH-942	650 ± 90	2000 ± 75
BPH-943	450 ± 110	2200 ± 150

Table 4.6: *In vivo* activity of bisphosphonates against *Plasmodium berghei* blood stage parasites.

Compound	**Dose (per mouse)**	**Pre-patent day (% parasitemia)**	**Delay in pre-patent period**
None	Untreated	1 (0.06% ± 0.01)	0
Risedronate	20 mg/kg for 5 days	2 (0.09%± 0.02)	1 day
BPH-715	1.5 mg/kg for 5 days	5 (0.01%± 0.003)	4 days

Mice were challenged on day 0 with $5x10^6$ parasites. Inhibitor treatment was given once a day on days -2, -1,0,+1,+2.

Table 4.7. Enzyme inhibition for *Plasmodium vivax* GGPPS, human FPPS and human GGPPS, together with SlogP values.

Compound	IC_{50}/ K_i for PvGGPPS (mM) [GPP][a]	IC_{50}/ K_i for PvGGPPS[a] (mM) [FPP]	IC_{50}/ K_i for HsFPPS (mM)[b]	IC_{50}/ K_i for HsGGPPS [c] (mM)	SlogP[d]
Risedronate	1.1 / 0.024	1.7 / 0.048	0.11 / 0.013	300 / 8.3	-5.5
BPH-715	0.56 / 0.013	1.5 / 0.043	0.032 / 0.0037	0.25 / 0.0069	-1.2
BPH-942	0.63 / 0.015	1.5 / 0.043	0.0041 / 0.0047	0.44 / 0.012	-4.1
BPH-943	0.70 / 0.016	1.5 / 0.043	0.016 / 0.0018	0.12 / 0.0033	-3.4

[a] Measured as described in Ref. 13.
[b] Measured as described in Ref. 16.
[c] Measured as described in Ref. 22.
[d] Calculated using the MOE program (Ref. 1).

Figure 4.1. Schematic illustration of isoprenoid biosynthesis pathways (in humans and *Plasmodium* spp) and sites of action of several inhibitors. Malaria parasites make IPP and DMAPP via the non-mevalonate pathway. Inhibition of isoprenoid biosynthesis affects protein prenylation; IPP build up activates γδ T cells; IPP is converted to toxic ATP analogs on FPPS inhibition. ANT, adenine nucleotide translocase; DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate, HMG, 3-hydroxy-3-methyl-glutaryl-CoA; TNF-α, tumor necrosis factor alpha; IFNγ, interferon gamma.

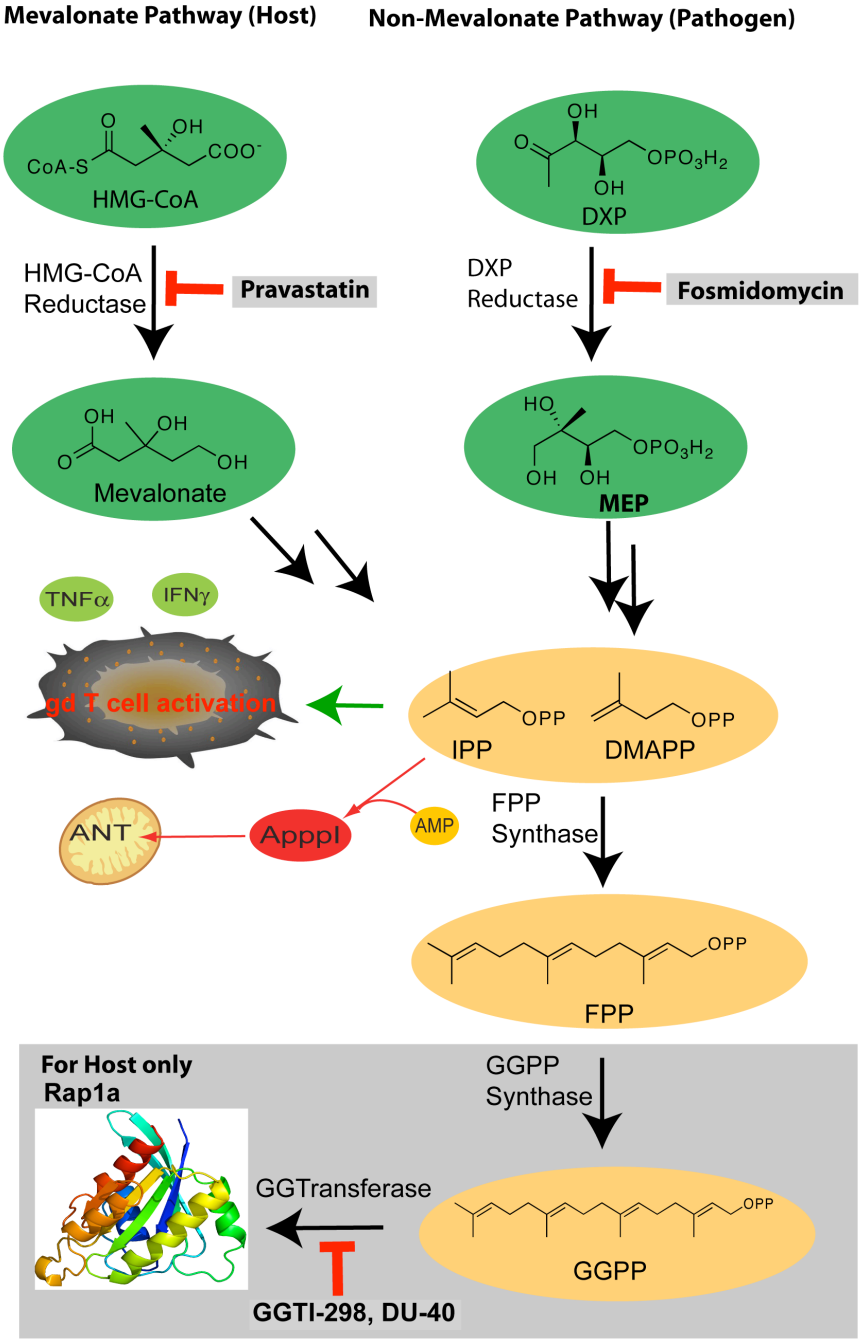


Figure 4.2. Structures of the inhibitors tested.

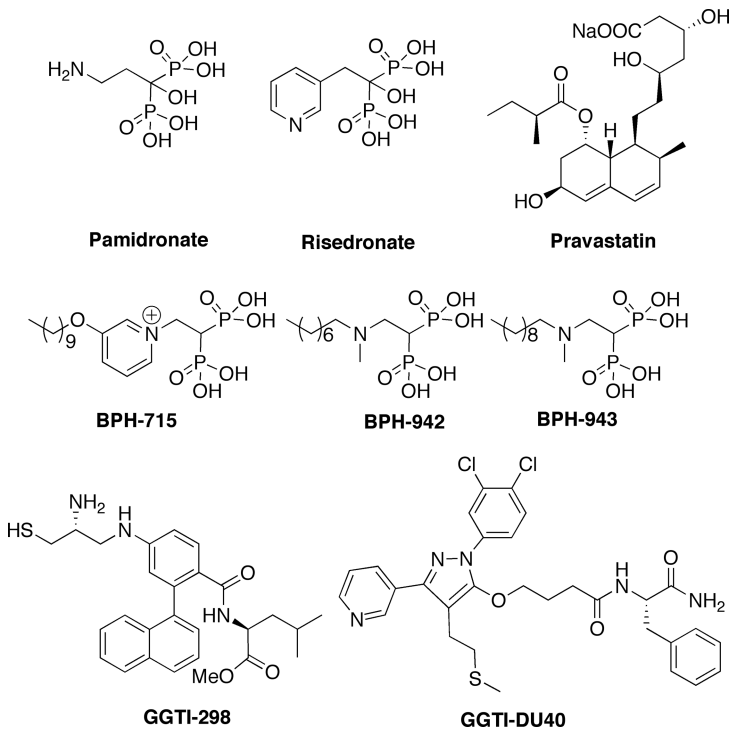


Figure 4.3. Typical dose-response curves for *in vitro* activity of two bisphosphonates *versus* *Plasmodium berghei* EEF cell growth inhibition.

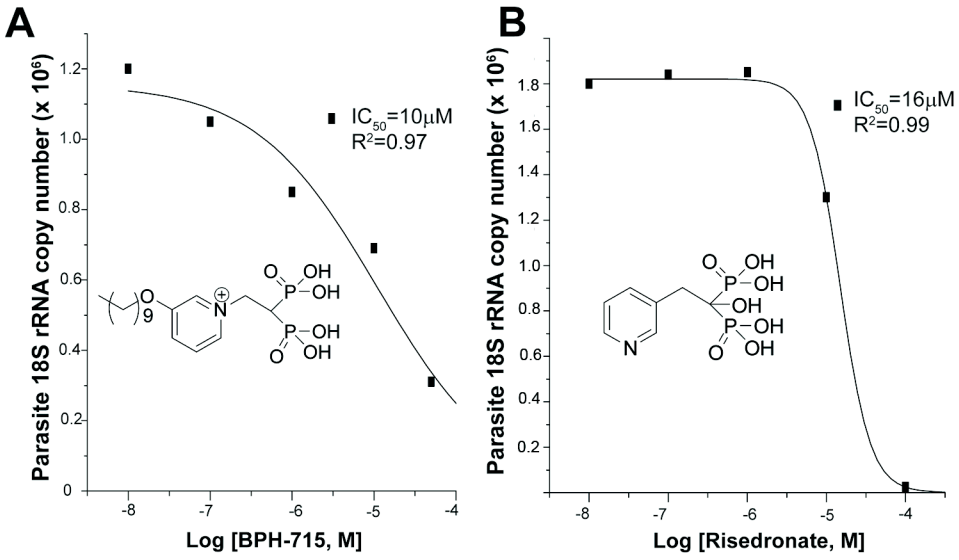
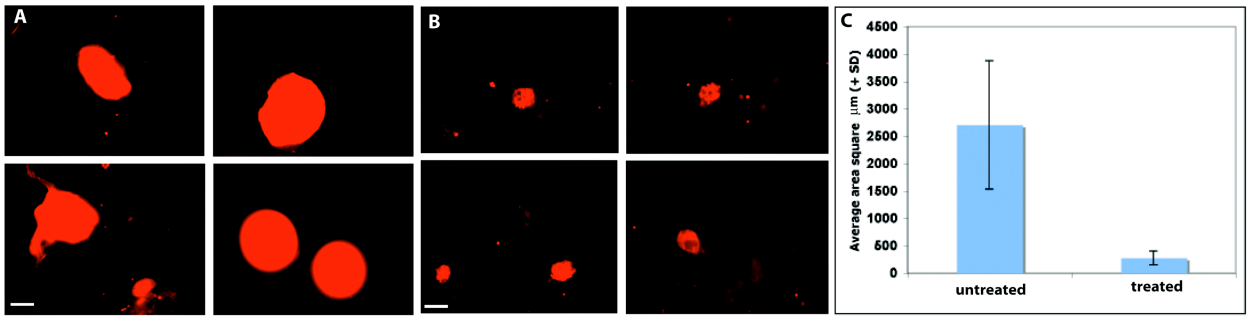


Figure 4.4. Results showing growth of untreated (A) and (B) treated (30 µM BPH-715) EEFs 50 hours post infection with sporozoites. EEFs were visualized by IFA using anti-HSP70 (parasite protein) monoclonal as the primary antibody and anti-mouse IgG Alexa 594 as the secondary antibody. BPH-715 treated EEFs show severe growth defects. (C), Area measurements of treated, and untreated, EEFs 50 hours post-infection with sporozoites.



4.7 References

1. Baird JK, *et al.* (2001) Randomized, parallel placebo-controlled trial of primaquine for malaria prophylaxis in Papua, Indonesia. *Clin Infect Dis* 33(12):1990-1997.
2. Valecha N, *et al.* (2001) Comparative antirelapse efficacy of CDRI compound 80/53 (Bulaquine) vs primaquine in double blind clinical trial. (Translated from English) *Current Science* 80(4):561-563 (in English).
3. Walsh DS, *et al.* (2004) Randomized trial of 3-dose regimens of tafenoquine (WR238605) versus low-dose primaquine for preventing *Plasmodium vivax* malaria relapse. *Clin Infect Dis* 39(8):1095-1103.
4. Dondorp AM, *et al.* (2009) Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 361(5):455-467.
5. Mutabingwa TK, *et al.* (2005) Amodiaquine alone, amodiaquine+sulfadoxine-pyrimethamine, amodiaquine+artesunate, and artemether-lumefantrine for outpatient treatment of malaria in Tanzanian children: a four-arm randomised effectiveness trial. *Lancet* 365(9469):1474-1480.
6. Mazier D, Renia L, & Snounou G (2009) A pre-emptive strike against malaria's stealthy hepatic forms. *Nat Rev Drug Discov* 8(11):854-864.
7. Wells TN, Alonso PL, & Gutteridge WE (2009) New medicines to improve control and contribute to the eradication of malaria. *Nat Rev Drug Discov* 8(11):879-891.
8. Singh AP, *et al.* (2007) Plasmodium circumsporozoite protein promotes the development of the liver stages of the parasite. *Cell* 131(3):492-504.
9. Rodan GA & Reszka AA (2002) Bisphosphonate mechanism of action. *Curr Mol Med* 2(6):571-577.

10. Kunzmann V, Bauer E, & Wilhelm M (1999) Gamma/delta T-cell stimulation by pamidronate. *N Engl J Med* 340(9):737-738.

11. Monkkonen H*, et al.* (2006) A new endogenous ATP analog (ApppI) inhibits the mitochondrial adenine nucleotide translocase (ANT) and is responsible for the apoptosis induced by nitrogen-containing bisphosphonates. (Translated from English) *British Journal of Pharmacology* 147(4):437-445 (in English).

12. Zhang YH*, et al.* (2009) Lipophilic Bisphosphonates as Dual Farnesyl/Geranylgeranyl Diphosphate Synthase Inhibitors: An X-ray and NMR Investigation. (Translated from English) *Journal of the American Chemical Society* 131(14):5153-5162 (in English).

13. Ghosh S*, et al.* (2004) Effects of bisphosphonates on the growth of *Entamoeba histolytica* and Plasmodium species *in vitro* and *in vivo*. (Translated from English) *Journal of Medicinal Chemistry* 47(1):175-187 (in English).

14. Martin MB*, et al.* (2001) Bisphosphonates inhibit the growth of *Trypanosoma brucei, Trypanosoma cruzi, Leishmania donovani, Toxoplasma gondii*, and *Plasmodium falciparum*: A potential route to chemotherapy. (Translated from English) *Journal of Medicinal Chemistry* 44(6):909-916 (in English).

15. Carraz M, Jossang A, Rasoanaivo P, Mazier D, & Frappier F (2008) Isolation and antimalarial activity of new morphinan alkaloids on *Plasmodium yoelii* liver stage. (Translated from English) *Bioorganic & Medicinal Chemistry* 16(11):6186-6192 (in English).

16. Zhang Y*, et al.* (Lipophilic pyridinium bisphosphonates: potent gammadelta T cell stimulators. *Angew Chem Int Ed Engl* 49(6):1136-1138.

17. Mukkamala D, No JH, Cass LM, Chang TK, & Oldfield E (2008) Bisphosphonate inhibition of a Plasmodium farnesyl diphosphate synthase and a general method for predicting cell-based activity from enzyme data. *J Med Chem* 51(24):7827-7833.

18. Sanders JM*, et al.* (2005) Pyridinium-1-yl bisphosphonates are potent inhibitors of farnesyl diphosphate synthase and bone resorption. *J Med Chem* 48(8):2957-2963.

19. Song Y*, et al.* (2009) Phosphonosulfonates are potent, selective inhibitors of dehydrosqualene synthase and staphyloxanthin biosynthesis in *Staphylococcus aureus*. *J Med Chem* 52(4):976-988.

20. C KMC*, et al.* (2008) Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates: a crystallographic and computational investigation. *J Med Chem* 51(18):5594-5607.

21. Bruna-Romero O*, et al.* (2001) Detection of malaria liver-stages in mice infected through the bite of a single Anopheles mosquito using a highly sensitive real-time PCR. *Int J Parasitol* 31(13):1499-1502.

22. Trager W & Jensen JB (1976) Human malaria parasites in continuous culture. *Science* 193(4254):673-675.

23. Mosmann T (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods* 65(1-2):55-63.

Chapter 5

Inhibition of the Fe_4S_4 Cluster-containing Protein IspH (LytB): EPR, Metallacycles and Mechanisms

5.1 Notes and Acknowledgements

This work was supported by the United States Public Health Service (NIH grants AI074233, GM073216, GM65307 and GM085774) and the Mississippi Center for Supercomputing Research. We thank Hassan Jomaa and Jochen Wiesner for providing their *A. aeolicus* IspH plasmid and for initial screening results, Pinghua Liu for the *E. coli* IspH plasmid, and Thomas B. Rauchfuss for helpful discussions.

5.2 Introduction

The Rohmer, methyl-D-erythritol phosphate or non-mevalonate pathway is responsible for isoprenoid biosynthesis in most pathogenic bacteria, as well as in malaria parasites.(1, 2) Since isoprenoids are essential for survival in these organisms and since the non-mevalonate pathway is absent in humans, enzymes that comprise the pathway are important anti-infective drug targets(3, 4) and in previous work (5, 6) it was shown that fosmidomycin (which inhibits the second enzyme in the pathway) gave very promising results in treating malaria. The structures and mechanisms of action of six of the eight enzymes present in the pathway have been known for some time, but deducing the structures (and mechanisms of action) of the last two enzymes has been more challenging. The penultimate enzyme is *E*-4-hydroxy-3-methyl-but-2-enyl diphosphate (HMBPP) synthase (IspG, GcpE, EC1.17.7.1) which catalyzes the $2H^+/2e^-$ reduction

of 2-*C*-methylerythritol-2,4-*cyclo*-diphosphate (MEcPP, **1**) to HMBPP (**2**), while the terminal enzyme, HMBPP reductase (IspH, LytB, EC1.17.1.2) catalyzes the $2H^+/2e^-$ reduction of HMBPP to isopentenyl diphosphate (IPP, **3**) and dimethylallyl diphosphate (DMAPP, **4**), in a ~5:1 ratio. The structure of IspG has not yet been reported, but there have been two structures published for IspH, one from *Aquifex aeolicus*,(7) the other from *E. coli*.(8) Both structures contain Fe_3S_4 clusters. However, based on the results of Mössbauer[9],[10] and EPR spectroscopy,(11) and from microchemical analyses,[9,] (11) we and others have concluded that Fe_4S_4 clusters are responsible for catalysis. The Fe_4S_4 cluster, while catalytically active, is thus relatively labile, as also found for example in aconitase and in pyruvate-formate lyase activating enzyme.(12) To date, there has been only one report of IspH inhibitors,(13) with IC_{50} values of ~1-2 mM having been obtained. Here, we report progress aimed at obtaining more potent inhibitors.

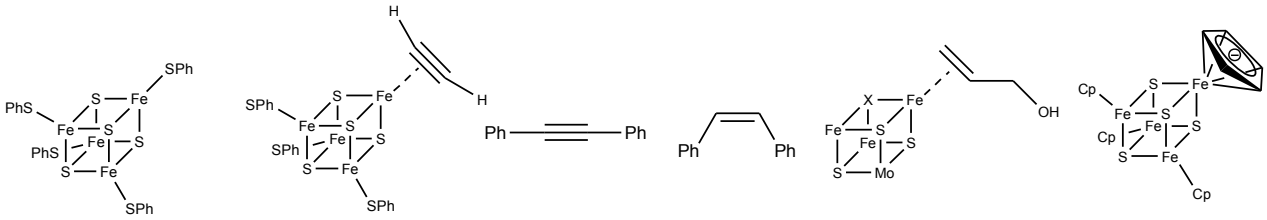
Based on the crystallographic, mutagenesis, catalytic activity and spectroscopic results described previously,(7-11) we reasoned that there might be two distinct approaches to developing IspH inhibitors. In the first, it might be possible to develop "conventional" inhibitors which bind to amino-acids involved in catalysis. In previous work(7) we proposed that a key residue involved in delivering H^+ to HMBPP was the totally conserved E126, and the essential nature of this residue in catalysis has now been confirmed by site-directed mutagenesis.[8] We reasoned that basic or cationic inhibitors might interact with this key residue, inhibiting the enzyme. The second hypothesis to test is that "unconventional" inhibitors, compounds that bind to the Fe/S cluster, might be developed. If the cluster is Fe_3S_4, this would be unprecedented and would thus, of course, not suggest what inhibitor structures to make. On the other hand, if the active species is Fe_4S_4 (as deduced from the IspH Mössbauer,(9, 10) EPR(11) and catalytic activity(9, 11) results), there is in fact precedent for the formation of organometallic species (i.e.

containing Fe-C bonds) between Fe_4S_4 clusters and alkynes, which would be expected to act as IspH inhibitors. For example, in the early literature, McMillan *et al.*(14) investigated the reduction of acetylene (C_2H_2) to ethylene (C_2H_4) by reduced Fe_4S_4 synthetic clusters, in particular $[Fe_4S_4(SPh)_4]^{3-}$, **5**. These workers proposed that an acetato complex reacted initially with C_2H_2 to form an organometallic species, **6**:

$$[Fe_4S_4(SPh)_4]^{3-} \xrightarrow{OAc^-} [Fe_4S_4(SPh)_3(OAc)]^{3-} + SPh^- \xrightarrow{C_2H_2} [Fe_4S_4(SPh)_3(C_2H_2)]^{2-} + OAc^-$$

5 **6**

most likely containing a side-on (π/σ) acetylene unit, which was then *cis*-reduced to ethylene. Basically the same reduction, of diphenylacetylene (**7**) to *cis*-stilbene (**8**), was reported by Itoh.(15)



5 **6** **7** **8** **9** **10**

In these systems, the alkyne complex was not observed directly, but under controlled potential (electrochemical reduction) conditions, Tanaka *et al.*(16) found evidence for a π complex of acetylene bound to $[Fe_4S_4(SPh)_4]^{3-}$ and $[Mo_2Fe_6S_8(SPh)_9]^{3-}$ clusters, as evidenced by quite large shifts in the C≡C vibrational Raman spectra. These workers also demonstrated that acetylene bound most strongly to reduced ($[Fe_4S_4]^+$) clusters, and resulted in release of 1 SPh^-. These results suggested to us that acetylenes might likewise bind to reduced IspH, forming organometallic (Fe-C) bonds. Interestingly, with low valent (Fe^I) complexes, alkynes have also been found to bind far more strongly than do alkenes,(17) leading again to the possibility that alkynes could bind to Fe and be good IspH inhibitors, displacing the olefinic substrate, HMBPP.

This formation of a π (or π/σ) "metallacycle" complex would be very similar to that deduced for the binding of the HMBPP parent molecule, allyl alcohol (which lacks the Me and CH_2OPP substituents) to the FeMo cofactor in nitrogenase,(18) illustrated schematically above (**9**) as the Fe_3MoS_3X cubane-like fragment. It also seemed possible that aromatic residues might interact with the Fe_4S_4 cluster, just as the cyclopentadienide ion does in model Fe_4S_4 clusters,(19) such as **10** (as an η^6 as opposed to an η^5 species). Although the structures of complexes such as **6** have not been confirmed crystallographically, the structure proposed by McMillan et al.(14) (**6**) is likely to involve the same type of bonding as found in many other organometallic complexes,(20) being described(18) as a resonance hybrid of a pure π complex and a σ complex, the latter corresponding to a metallacycle.

5.3 Results and Discussion

In an effort to learn more about just what types of compound might bind to the Fe_4S_4 cluster in IspH, we first screened a series of small molecules, using EPR spectroscopy. The 9 GHz EPR spectrum of *A. aeolicus* IspH in the g ~ 2 region exhibits a broad spectrum (Figure 5.1a), very similar to that of the *E. coli* protein (Figure 5.1b) and is characteristic of an S=1/2 $[Fe_4S_4]^+$ species.(11) The unliganded *A. aeolicus* IspH also has two small signals at g~5 (Figure 5.2a), which likely indicate the presence of higher spin species. For purpose of comparison, the g-values for the systems studied here are presented in Table 5.1, together with g-values for related systems such as aconitase, and IspG (GcpE).

We first tested a series of small molecules and ions: CO, N_3^-, MeCN and CN^-, for cluster-binding in *A. aeolicus* IspH, but only CN^- was found to have any effect on the EPR spectrum, Figure 5.1c (Figure 5.2b).The g-values observed were 2.08/2.05 and 1.94, Figure 5.1c, close to

those observed previously with CN^- binding to the *Pyrococcus furiosus* 4Fe ferredoxin (g_1 = 2.09, g_2 = 1.95, and g_3 = 1.92), another 4Fe-4S cluster with 3 Cys ligands and a unique 4th Fe position,(21) due we propose to end-on binding to the 4th Fe in IspH. There was also an increase in signal intensity in the g~2 region (from ~20% to ~60% spin/protein), due to conversion of the higher spin state species to S=1/2.

We then tested a series of other small molecules that might be anticipated to bind in a sideways-on mode (as with proposed for acetylene in the model systems): propargyl alcohol, propargylamine and propiolic acid. All species resulted in large spectral changes, and by way of example the IspH + propiolic acid ($HC{\equiv}CCO_2H$) spectrum is shown in Figure 5.1d. These results indicated to us that, as with the Fe_4S_4 containing small molecules, acetylenes bind to IspH, forming π or π/σ complexes. With both CN- as well as propiolic acid binding there also appear to be two bound ligand conformations or binding modes (e.g. hydrogen bonded or not), as evidenced by the splitting of the g_1 signals.

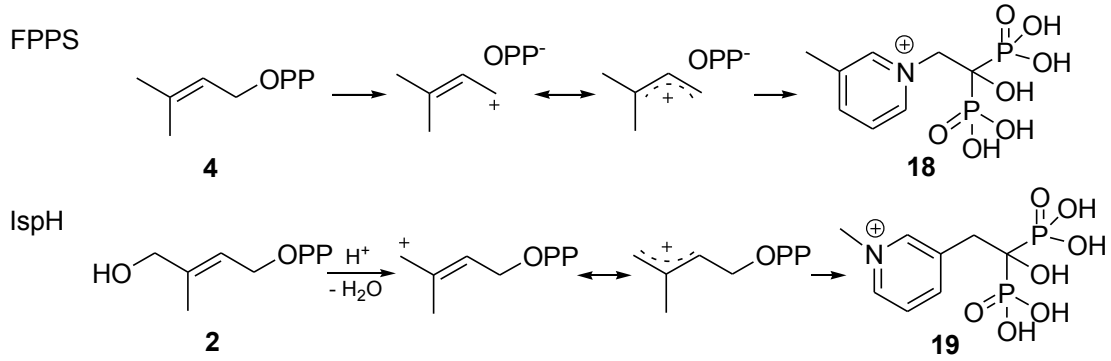
However, propiolic acid had poor activity in IspH inhibition, and we reasoned that acetylenic diphosphates (or the isoelectronic, cyano diphosphates) might interact more strongly with IspH, since there would be an increase in binding affinity due to their diphosphate moieties docking into the (inorganic) PPi site seen crystallographically.(8) We thus synthesized four acetylenic diphosphates (**11** - **14**, Figure 5.3) and two cyanoalkyl diphosphates (**15**, **16**) and tested them for their activity in IspH inhibition. IC_{50} results are shown in Figure 5.3. The most active compound was **12** with an IC_{50} = 0.45 μM (K_i ~ 60 nM; Figure 5.4a). As shown in Figure 5.4b, this compound can bind with its diphosphate group in the inorganic "PPi" site seen crystallographically,(8) with the alkyne fragment interacting with the unique 4th Fe (identified from EPR, Mössbauer and catalytic activity results, and added computationally here as

described previously(7)), with Fe-C bond lengths of ~2.4 Å, suggesting π-complex formation. There was a large change in the EPR spectrum of IspH in the presence of **12** (Figure 5.1e, Figure 5.2c) indicating a major change in the cluster's electronic structure, consistent with these docking results. Moreover, there was evidence for only a single bound species. The shorter chain analog **11** was less active than was **12**, as was the longer chain analog **13**, and conversion of the bridging O to CH_2 (**17**) reduced the activity of **11** even more (Figure 5.3), due presumably to a disruption in H-bonding to the SXN motif. Likewise, the addition of a terminal CH_2OH (**14**) reduced activity, consistent with the known stronger binding of terminal versus substituted alkynes to low valent Fe complexes.(17) The changes in g-values on addition of the acetylenes were generally similar, as expected, and as with CN^- there was a ~3x increase in signal intensity in the g~2 region (due to a high-to-low spin state conversion).We also tested the effects of the isoelectronic analogs (**15**, **16**) of the acetylenes (-C≡C-H → -C≡N), cyanides, on IspH inhibition, but both compounds were far less active than their acetylenic counterparts.

These results clearly indicate that alkyne diphosphates can be good IspH inhibitors, with the most active species being > 1,000 x more potent than previously reported IspH inhibitors.(13) Since PPi itself and other diphosphates are only weak (~1 mM) IspH inhibitors,(8) we conclude that ligand binding is driven by π (or π/σ) or η^2-alkynyl complex formation with the Fe_4S_4 cluster, the same type of complex formation as suggested by Raman spectroscopy,(16) for acetylene binding to an $[Fe_4S_4(SPh)_4]^{3-}$ cluster. Moreover, the results of density functional theory calculations show that there is indeed Fe-C metal ligand bonding present in an $[Fe_4S_4(SMe)_3(HC{\equiv}CCH_2OH)]^{2-}$ model system. We show in Figures 5.4 c,d some typical MO results. The αHOMO-1 shows the interaction of an Fe d orbital with the C≡C π orbital, in the plane of the Fe…C≡C fragment. The βHOMO-1 illustrates the interaction between an Fe d

orbital with another type of C≡C π orbital, this time perpendicular to the plane of Fe…C≡C. If IspH, under these reducing conditions, contained an Fe_3S_4 cluster, there would be no obvious way in which acetylenes would bind tightly to the cluster (or the protein), while η^2-alkynyl (i.e. π or π/σ) complex formation is predicted, based on previous work, and is supported by the DFT results.

We next investigated a series of cationic (or basic) diphosphates and (cationic) bisphosphonates, which are isosteres of the diphosphate group. In previous work(22) we found that aromatic, cationic bisphosphonates (such as **18**) were potent (K_i ~ 20 nM) inhibitors of the enzyme farnesyl diphosphate synthase since, as shown below:



their charge distribution mimics that expected for the allylic cation / diphosphate anion ion-pair.(23) In the case of IspH, similar allylic cations (shown above) have been proposed(24) to be involved in catalysis, so we tested both **18** and **19** for their activities in IspH inhibition. Neither had potent activity (**18**, IC_{50} = 540 μM; **19**, IC_{50} > 1mM, Figure 5.3). We then tested six additional bisphosphonates (**20** - **25**, Figure 5.3) with 1-H, 1-Me or 1-OH backbone groups and pyridine, pyridinium or amino-pyridine side-chains. The most active compound (IC_{50} = 67 μM) was **21**, an amino-pyridine, expected to contain an amidinium-like (protonated) side-chain. Overall, however, the activity of these bisphosphonates was only modest, and based on Glide

docking results using the "closed" IspH structure (PDB File #3F7T) it appeared that this might be due to the "branched" nature of the bisphosphonate backbone. We thus next synthesized a series of pyridinium diphosphates, **26** - **29**, which based on computational docking results using Glide(25) appeared to better fit the IspH active site. The most active compound in the series (Figures 5.3, 5a) was **26**, a *meta* pyridinium diphosphate having an IC_{50} = 38 μM, slightly more potent than the best bisphosphonate, **21**. When docked into the IspH active site, we find (Figures 5.5b, c) that when the diphosphate backbone binds to the "PPi" site seen crystallographically, the pyridinium H^N is only ~1.9 Å from the E126 O (modeled as CO_2^-), indicating the possibility of an H-bond or electrostatic interaction with this active site residue. The *para*-pyridinium analog (**27**) was less active (IC_{50} = 149 μM), and the *ortho*-pyridinium analog (**28**) was far less active (IC_{50} = 1.2 mM), correlating with the increased NH-O (CO_2^-) distances of 3.47 Å and 4.28 Å, Figure 5.5c. Interestingly, all three aromatic rings also locate close to the 4th Fe, but the major decrease in activity seen as the ring nitrogens move away from the E126 carboxyl indicates the dominance of a Coulombic interaction, consistent with the lack of activity of the chloropyridine analog, **29,** which is expected to be far less basic than is **26**. These activity and docking results are consistent, then, with the pyridinium diphosphates binding to IspH with their PPi groups binding to the inorganic PPi site, while their aromatic rings interact with E126, as illustrated in Figures 5.4b, c. There was no major change in the EPR spectrum of **26** bound to the protein, Figure 5.2d, consistent with the lack of any π-complex interaction. We then sought to improve activity by varying the length of the CH_2 spacer in the side-chain: addition of 1 CH_2 group (**30**) resulted in an IC_{50} = 9.1 μM (K_i ~ 1.2 μM), but addition of 2 CH_2 groups reduced activity (**31**, IC_{50} = 463 μM; Figure 5.3), with the pyridinium nitrogen in **30** again docking close to E126, while **31** did not. Taken together, these results argue against formation of a carbocation

mechanism since for this, we would anticipate good activity for **28** (with maximum charge density separated by two carbons from the PPi backbone), while experimentally, maximum activity is found with **30** (with its NH^+ located four carbons from the PPi backbone). That is, none of the cationic (**18**, **19**) or pyridine/pyridinium analogs of the putative carbocation transition states (or reactive intermediates) act as inhibitors, while the longer chain species that can bind to both the PPi site as well as E126, are the most potent inhibitors among the whole series of compounds (**18-31**) investigated. Notably, with both the alkyne as well as the cationic inhibitors, the side-chains are located (Figure 5.5d) close to the Fe_4S_4 cluster blocking, we propose, HMBPP access.

These results lead to the proposals for IspH inhibition shown in Figure 5.6. We propose that both alkyne as well as alkene diphosphates (i.e. HMBPP) bind to the Fe_4S_4 cluster at the unique 4th Fe, forming π (or π/σ) "metallacycles", as shown in Figures 5.6a, b, the same type of binding reported previously for the HMBPP parent molecule (allyl alcohol) bound to the FeMo cofactor in nitrogenase.(18) As shown in Figure 5.6b, the presence of such a metallacycle also immediately leads to a proposal for the IspH catalytic mechanism in which in a first H^+/e^- step, the HMBPP 4-OH is protonated and an η^1-allyl intermediate and water form. This intermediate is then cleaved in a second H^+/e^- step, forming the DMAPP product, Figure 5.6b. IPP forms in the same way, after an η^1-allyl (or η^3-allyl) isomerization. This model is described in more detail elsewhere(26) and is supported by electron nuclear double resonance spectroscopy using [u-^{13}C] and [^{2}H]-labeled HMBPP, which strongly suggest metal-ligand interaction. We thus propose that the alkyne as well as the cationic inhibitors block or displace HMBPP from binding to the Fe_4S_4 cluster: the alkynes form π (or π/σ) "η^2-alkynyl" metallacycles, while the cationic inhibitors interact with the active site E126 carboxylate, again effectively blocking HMBPP binding.

Validation of the Mechanism of IspH Inhibition, and Catalysis.

Finally, after review of this article, Grawert et al.(27) reported the results of new crystallographic structures of HMBPP, PPi, IPP and DMAPP bound to IspH(27) which help us validate the inhibition models proposed here. These workers now report that HMBPP binds as an alkoxide to the unique 4th iron (to the oxidized $[Fe_4S_4]^{2+}$ cluster), the same binding mechanism we proposed earlier.(7) On irradiation in the x-ray beam, HMBPP is reduced, forming a deoxy species (bound to the Fe_4S_4 cluster), basically as we proposed based on EPR/ENDOR and docking results.(26) These x-ray results are of interest since they enable us to make a direct comparison between the x-ray and EPR/ENDOR/docking based methods,(26, 27) helping us validate the use of inhibitor docking. We show in Figure 5.7a an alignment of the x-ray structure of HMBPP bound to IspH with that we proposed (Figure 5.1G in ref (26)) for the initial HMBPP binding pose in which the O-4 can interact with the unique 4th Fe. There is a 0.32 Å rmsd between the heavy atoms (C, O, P) in the HMBPP ligand between the x-ray and our docking/EPR based structures, ~ 0.43 Å if we include the Fe, S atoms in the cluster, Figure 5.7a. On reduction, we proposed(26) that the 4-OH is removed from the ligand and the σ/π metallacycle is reduced, resulting in loss of water and formation of a η^1/η^3-allyl complexes and thence, IPP and DMAPP.

The new crystallographic results are thus in good accord with those we proposed from our EPR, ENDOR and computational work, which gives confidence in the proposed ligand geometries we describe here for the bound inhibitors. As can be seen in Figure 5.7b, the alkyne group in propargyl diphosphate (**11**) bound to IspH(26) is located in essentially the same position as the deoxy product species seen crystallographocally, where it can interact with the unique, 4th iron, forming a π or π/σ complex. The alkynes are much stronger inhibitors of IspH than are the alkenes (IPP, DMAPP),(8) and indeed the ENDOR spectrum of **11** bound to IspH

exhibits an ~6 MHz ^{13}C hyperfine interaction,(26) 2-3 times larger than the ^{13}C hyperfine interactions found for either allyl alcohol bound to nitrogenase,(28) or HMBPP bound to IspH.(26)

5.4 Conclusions

Taken together, these results are of interest for several reasons. First, we find that cationic diphosphates are IspH inhibitors. However, rather than mimicking transition states /reactive intermediates, we propose that they interact with the active site E126, preventing HMBPP binding to the Fe_4S_4 cluster. Second, we find that alkyne diphosphates are quite potent (IC_{50} ~ 0.45 μM; K_i ~ 60 nM) IspH inhibitors. This activity was predicted based on previous Raman and catalytic activity results with model systems, which have been interpreted as indicating π (or π/σ) "metallacycle" formation between Fe_4S_4 clusters and acetylene. That alkynes are, in fact, IspH inhibitors, also supports the idea that catalytically active IspH contains active-site Fe_4S_4 and not Fe_3S_4 clusters, since there is no obvious way in which an alkyne group would bind to the three exposed S^{2-} in such clusters. On the other hand, there is good literature precedent for alkynes binding to Fe_4S_4 clusters, supporting formation in IspH of π (or π/σ) complex formation. Third, the results of quantum chemical calculations indicate Fe-C metal ligand bonding with the acetylene group (in a $[Fe_4S_4(SMe)_3(HC\equiv CCH_2OH)]^{2-}$ model. Fourth, the formation of acetylene metallacycles is consistent with the observation that the HMBPP parent molecule, allyl alcohol, forms a metallacycle with the nitrogenase FeMo cofactor, which contains a related Fe/S (Mo, "X") cluster and where the ligand forms Fe – C bonds with a single Fe. The observation that the IC_{50} for the most potent IspH alkyne inhibitor (IC_{50} = 0.45 μM; K_i ~ 60 nM) is much smaller than the K_M for HMBPP binding (K_M = 5 μM) is also expected based on organometallic precedent in

which alkynes bind much more strongly (to low valent Fe complexes) than do alkenes.(17) Fifth, we find that CN^- binds to IspH with the resulting spectrum quite strongly resembling that found with *P. furiosus* Fd, another protein with an unique 4th Fe. And finally, the recent revision of the earlier 3 Fe IspH structural/mechanistic models is in accord with our earlier mechanistic proposal of initial binding of the HMBPP 4-OH (as an alkoxide) to an $[Fe_4S_4]^{2+}$ cluster which, on reduction, forms an allyl complex which is then cleaved to form the IPP and DMAPP products.(26) This gives confidence in the models we propose here for alkyne inhibition. These results also open up the intriguing possibility that metallacycle complexes may act as reactive intermediates (or transition states) in both IspG as well as IspH catalysis, and that alkynes can be expected to be good inhibitors of other Fe_4S_4 cluster-containing enzymes which contains "unique" 4th Fe atoms.

5.5 Materials and Methods

A. aeolicus IspH protein purification.

BL-21(DE3) cells expressing IspH from *A. aeolicus* were grown in LB media supplemented with 150 mg/mL ampicillin at 37 °C until the OD_{600} reached 0.6. Cells were then induced with 200 μg / L anhydrotetracycline, then grown at 20 °C for 15 hours. Cells were harvested by centrifugation (9000 rpm, 8 min, 4 °C) and were kept at -80 °C until use. Cell pellets were resuspended and lysed in B-PER (Thermo Scientific, Rockford, IL) protein extraction reagent for one hour at 4°C, then centrifuged at 200,000 rpm at 4 °C for 15 min. The supernatant was applied to a Ni-NTA column equilibrated with 5 mM imidazole in pH 8.0 buffer containing 50 mM Tris·HCl and 150 mM NaCl. After washing with 20 mM imidazole, protein was eluted with 100 mM imidazole. Fractions were collected and dialyzed in pH 8.0 buffer containing 50 mM Tris·HCl, 150 mM

NaCl, 5% glycerol, and 1 mM DTT, 4 times. The purified protein was flash-frozen in liquid nitrogen and stored at -80 °C until use.

***E. coli* IspH Protein Purification.**

BL21 DE3 (Invitrogen) cells harboring an *E. coli* IspH construct were grown in LB media at 37°C until the OD_{600} reached 0.6. Induction was performed with 200 ng/ml anhydrotetracycline at 20°C for 15 hours. Cells were harvested by centrifugation at 9000 rpm for 8 mins and stored at -80 °C. Cell pellets were resuspended and lysed in B-PER (Thermo Scientific, Rockford, IL) protein extraction reagent for about one hour at 4°C, then the lysate was centrifuged at 250,000 rpm for 30 mins. The supernatant was collected and loaded onto an IBA Strep-tag column equilibrated with buffer W (100 mM Tris·HCl, 150 mM NaCl, pH 8.0). After washing with buffer W, protein was eluted using buffer E (buffer W containing 2.5 mM desthiobiotin). Fractions were collected and dialyzed in pH 8.0 buffer containing 50 mM Tris·HCl, 150 mM NaCl, 5% glycerol, and 1mM DTT, twice. The purified protein was flash-frozen in liquid nitrogen and stored at -80 °C until use.

Protein reconstitution.

Both *A. aeolicus* and *E. coli* IspH proteins as isolated had a very small peak at 410 nm (A_{280}/A_{410} < 0.02), so were reconstituted for further studies. Before reconstitution, protein was transferred into a Coy Vinyl Anaerobic Chamber after being degassed on a Schlenk line. The following steps were performed inside the anaerobic chamber with an oxygen level < 2 ppm. In a typical reconstitution experiment, 10 mM DTT and ~ 0.5 mg of elemental sulfur were added to 3 mL 0.6 mM protein solution in a pH 8.0 buffer containing 50 mM Tris·HCl, 150 mM NaCl and 5% glycerol. After stirring for 1.5 hours, $FeCl_3$ was slowly added from a 30 mM stock solution to 6 equivalents. After 3 hours, an aliquot of the solution was centrifuged and a UV-VIS spectrum

recorded. If the A_{410nm} / A_{280nm} ratio was ≥ 0.38, the protein was then desalted by passing through a PD10 column. If the ratio was < 0.38, more DTT, elemental sulfur and $FeCl_3$ were added and the sample incubated with stirring (for typically ~ 2 hours) until the 410 nm / 280 nm absorption ratio was ~ 0.38. The reconstituted protein was then concentrated by ultrafiltration, and the protein concentration determined by using a Bio-Rad (Hercules, CA) Protein Assay kit.

Enzyme inhibition assays.

All assays were performed anaerobically at room temperature according to Altincicek et. al.(24) with minor modification. To a pH 8.0 buffer solution containing 50 mM Tris·HCl, 150 mM NaCl, and 5% glycerol, sodium dithionite was added to 0.4 mM, methyl viologen was added to 2 mM, and IspH was added to 72 nM. For enzyme assays, various amounts HMBPP were added and the reactions were monitored at 732 nm. The initial velocities were fitted by using the Michaelis-Menten equation with OriginPro 8 (OriginLab Corporation, Northampton, MA) software. The activity of reconstituted *A. aeolicus* IspH tested under the conditions described above was 1.2 μmol min^{-1} mg−1 with a K_M=7 μM. For inhibition assays, various concentrations of inhibitor were added and incubated for 10 min, prior to addition of 34 μM HMBPP. Initial velocities at different inhibitor concentrations were then plotted as dose-response curves, and were fitted to the following equation, from which IC_{50} values were determined:

$$y = \frac{1}{1 + \left(\frac{x}{IC_{50}} \right)^{slope}}$$

where x is the inhibitor concentration and y is the fraction inhibition. K_i values were then deduced from the IC_{50} value by using the Cheng-Prusoff equation:(29)

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_M}}$$

where [S] is the HMBPP concentration, and K_M is the Michaelis constant.

EPR spectroscopy.

Samples for EPR spectroscopy were typically 0.3 mM in IspH, and were reduced by adding 20 equivalents of sodium dithionite followed by incubating for 5 min. Glycerol was added to 42.5 % (v/v). EPR samples were frozen in liquid nitrogen after reduction. EPR spectra were collected at X-band using a Varian E-122 spectrometer together with an Air Products (Allentown, PA) helium cryostat. Data acquisition parameters were typically: field center = 3250 G; field sweep = 800 G; modulation = 100 kHz; modulation amplitude = 5 G; time constant = 32 ms; 60 sec per scan; 8 sec between each scan; and temperature = 15K. EPR spectral simulations were carried out by using the EasySpin program.(30)

Docking calculations.

For docking calculations, the IspH target protein (PDB: 3F7T) was prepared using the protein preparation wizard in Maestro 8.0.(31) Water from the active site region was removed, as was the diphosphate ligand. The Fe_3S_4 cluster was reconstituted computationally to form the Fe_4S_4 species as described previously(7) and hydrogen atoms were added to the protein. Hydrogen bonds were optimized to default values and an energy minimization in Macro-Model 9.5[32] performed only on the protein hydrogens, using default parameters. A receptor grid large enough to encompass all crystallographically-observed binding sites was then generated from the prepared target protein. Geometry optimized ligands were docked using Glide(25) extra-precision (XP) mode, and no other constraints were applied. In some instances, we also used the MMFF94 force-field(33) to effect further geometry optimization.

Density functional theory calculations.

In order to gain a better understanding of the interaction between the propargyl diphosphate inhibitors and the Fe-S cluster, we used the published structure of the lowest energy form of allyl alcohol bound to the nitrogenase FeMo cofactor (structure 3 in ref.18), converting Mo → Fe, X → S and allyl → alkynyl as the initial structure. Geometry optimization was performed by using the pure density functional theory (DFT) method with a BPW91 functional,(34, 35) a Wachter's basis (62111111/3311111/3111) for Fe,(36) 6-311G* for all the other heavy atoms, and 6-31G* for the hydrogens, using the Gaussian 09 program.(37) This method is similar to that used in the calculations of the ligand-bound nitrogenase structures(18) and is the same as that we used previously to make accurate predictions of NMR hyperfine shifts, ESR hyperfine couplings, as well as Mössbauer quadrupole splittings and isomer shifts, in various iron-containing proteins and model systems.(38-43)

Synthetic aspects.

General methods All reagents used were purchased from Aldrich (Milwaukee, WI). The purities of all compounds investigated were confirmed either via combustion analysis (for solid samples) or by using ^{1}H and ^{31}P NMR spectroscopy at 400 or 500 MHz on Varian (Palo Alto, CA) Unity spectrometers. Cellulose TLC plates were visualized by using iodine or a sulfosalicylic acid-ferric chloride stain.(44) Propargyl methanesulfonate and but-3-ynyl methanesulfonate were synthesized according to a literature method.(45, 46) The syntheses and characterization of compounds **18 - 25** have been described previously.(22, 47-51) and the samples used here were from the same batches as described.

General procedure for preparation of diphosphates Diphosphates were prepared using modified literature procedures.(52) Typically, 0.5-1 mmol of halide or mesylate in a minimum

amount of CH_3CN (0.4-0.6 mL) was added dropwise to a stirred solution of 2-3 equiv of tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3-6 mL) at 0°C, then the reaction mixture was allowed to stir for 2-24 h at room temperature and solvent removed under reduced pressure. The residue was dissolved in cation-exchange buffer (49:1 (v/v) 25mM NH_4HCO_3/2-propanol) and slowly passed over 60-100 mequiv Dowex AG50W-X8 (100-200 mesh, ammonium form) cation-exchange resin, pre-equilibrated with two column volumes of the same buffer. The product was eluted with two column volumes of the same buffer, flash frozen, then lyophilized. The resulting powder was dissolved in 50 mM NH_4HCO_3. 2-Propanol/CH_3CN (1: l (v/v)) was added, and the mixture vortexed, then centrifuged for 5 min at 2000 rpm. The supernatant was decanted. This procedure was repeated three times, and the supernatants were combined. After removal of the solvent, lyophilization, then flash chromatography on a cellulose column, a white solid was obtained.

Synthesis of individual compounds

Prop-2-ynyl diphosphate (11). Propargyl methanesulfonate (134 mg, 1 mmol) in CH_3CN (0.5 mL) was added dropwise to a stirred solution of 2.70 g (3.0 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL) at -20°C. The reaction mixture was then slowly warmed to room temperature over 2 h and solvent removed under reduced pressure. Flash chromatography on a cellulose column (4:1:2.4 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 47 mg (18 %) of a white solid. ^{1}H NMR (400 MHz, D_2O) δ 2.68 (s, 1H), 4.38 (d, $J_{H,P}$ = 9.2 Hz, 2H); ^{31}P NMR (162 MHz, D_2O) δ -10.10 (d, J = 20.7 Hz), -7.67 (d, J = 20.7 Hz).

But-3-ynyl diphosphate (12). But-3-ynyl methanesulfonate (148 mg, 1 mmol) in CH_3CN (0.5 mL) was added dropwise to a stirred solution of 2.70 g (3.0 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL) at 0°C. The reaction mixture was allowed to warm to

room temperature over 24 h and solvent removed under reduced pressure. Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 28 mg (10%) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 2.16-2.17 (m, 1H), 2.35-2.40 (m, 2H), 3.81-3.84 (m, 2H); ^{31}P NMR (162 MHz, D_2O): δ -9.83 (d, J = 17.0 Hz), -7.82 (d, J = 15.9 Hz).

Pent-4-ynyl diphosphate (13). Pent-4-ynyl methanesulfonate (162 mg, 1 mmol) was treated with 2.70 g (3 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 103 mg (35%) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 1.65-1.70 (m, 2H), 2.13-2.17 (m, 3H), 3.83 (q, J = 6.8Hz, 2H). ^{31}P NMR (162 MHz, D_2O): δ -9.64 (d, J = 20.9 Hz), -7.76 (d, J = 20.7 Hz).

4-Hydroxybut-2-ynyl diphosphate (14). 4-Chlorobut-2-yn-1-ol (104 mg, 1 mmol) was treated with 2.70g (3 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded a white solid. ^{1}H NMR (400 MHz, D_2O): δ 4.07(s, 2H), 4.31(d, $J_{H,P}$ = 6.8Hz, 2H); ^{31}P NMR (162 MHz, D_2O): δ -10.11(d, J = 20.8 Hz), -8.85 (d, J = 20.7 Hz).

Cyanomethyl diphosphate (15). 2-Chloroacetonitrile (76 mg, 1 mmol) was treated with 2.70 g (3 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 45 mg (17%) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 4.56 (d, J = 10.4 Hz, 2H); ^{31}P NMR (162 MHz, D_2O): δ -10.25 (d, J = 22.0 Hz), -7.37 (d, J = 20.7 Hz).

2-Cyanoethyl diphosphate (16). 3-Bromopropanenitrile (134 mg, 1 mmol) was treated with 1.80 g (2 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (4 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3)

yielded 31 mg (11 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 2.68 (t, J = 6.4 Hz, 2H), 3.96 (q, J = 6.4 Hz, 2H). ^{31}P NMR(162 MHz, D_2O): δ -10.24 (d, J = 20.7 Hz), -6.21 (d, J = 20.7 Hz).

[[(Prop-2-ynyl) phosphinyl] methyl] phosphonic acid (17). Propargyl methanesulfonate (134 mg, 1 mmol) was treated with 2.70 g (3 mmol) tris(tetra-n-butylammonium) hydrogen methanediphosphonate in CH_3CN (4 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 93 mg (35 %) of a white solid. ^{1}H NMR (500 MHz, D_2O): δ 2.00 (t, $J_{H,P}$ = 12.0 Hz, 2H), 2.70 (t, J = 2.0 Hz, 1H), 4.37 (dd, $J_{H,P}$ = 9.2 Hz, J = 2.0 Hz, 2H); ^{31}P NMR (202 MHz, D_2O): δ 15.59 (d, J = 9.3 Hz), 20.06 (d, J = 10.7 Hz).

(Pyridin-3-yl)-methyl-diphosphate (26). 3-(Bromomethyl)pyridine (86 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3 mL). Flash chromatography on a cellulose column (3:2 (v/v) 2-propanol/50mM NH_4HCO_3) yielded 73 mg (45 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 4.80 (d, $J_{H,P}$ = 7.2 Hz, 2H), 7.31(dd, J = 7.6 Hz, J = 4.8 Hz, 1H), 7.81 (d, J = 7.6 Hz, 1H), 8.30 (d, J = 4.8 Hz, 1H), 8.42 (s, 1H); ^{31}P NMR (162 MHz, D_2O): δ -9.74 (d, J = 22.0 Hz), -5.95 (d, J = 22.0 Hz).

(Pyridin-4-yl)-methyl-diphosphate (27). 4-(Bromomethyl)pyridine (86 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3 mL). Flash chromatography on a cellulose column (3:2 (v/v) 2-propanol/50mM NH_4HCO_3) yielded 65 mg (40 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 4.92 (d, $J_{H,P}$ = 8.0 Hz, 2H), 7.41 (d, J = 6.0 Hz, 2H), 8.34 (d, J = 5.6 Hz, 2H); ^{31}P NMR (162 MHz, D_2O): δ -9.66 (d, J = 20.7 Hz), -6.49 (d, J = 20.7 Hz).

(Pyridin-2-yl)-methyl-diphosphate (28). 2-(Bromomethyl)pyridine (86 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3

mL). Flash chromatography on a cellulose column (3:2 (v/v) 2-propanol/50 mM NH_4HCO_3) yielded 56 mg (35%) of a white solid. ^{1}H NMR(400 MHz, D_2O): δ 4.91 (d, $J_{H,P}$ = 7.2 Hz, 2H), 7.28 (t, J = 4.8 Hz, 1H), 7.49 (d, J = 8.0 Hz, 1H), 7.80 (t, J = 8.0 Hz, 1H), 8.28 (d, J = 4 Hz, 1H); ^{31}P NMR (162 MHz, D_2O): δ -9.55 (d, J = 22.0 Hz), -6.02 (d, J = 19.4 Hz).

(6-Chloropyridin-3-yl)-methyl-diphosphate (29). 5-(Bromomethyl)-2-chloropyridine (103 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3 mL). Flash chromatography on a cellulose column (3:2 (v/v) 2-propanol/50 mM NH_4HCO_3) yielded 70 mg (40 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 4.82 (d, $J_{H,P}$ = 7.6 Hz, 2H), 7.29 (d, J = 8.4 Hz, 1H), 7.74 (dd, J = 8.4 Hz, J = 2.4 Hz, 1H), 8.20 (d, J = 2.4 Hz, 1H). ^{31}P NMR (162 MHz, D_2O) : δ -9.73 (d, J = 22.0 Hz), -5.81 (d, J = 22.0 Hz).

Pyridin-3-yl-ethyl-diphosphate (30). 3-(2-Bromoethyl) pyridine (93 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3 mL). Flash chromatography on a cellulose column (2:1:1 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 42 mg (25 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 2.87 (t, J = 6.4 Hz, 2H), 4.00 (q, J = 6.4 Hz, 2H), 7.33 (t, J = 6.4 Hz, 1H), 7.79 (d, J = 8.0 Hz, 1H), 8.24 (s, 1H), 8.36 (s, 1H); ^{31}P NMR (162 MHz, D_2O): δ -9.78 (1 d, J = 20.6 Hz), -7.07 (d, J = 20.6 Hz).

Pyridin-3-yl-propyl-diphosphate (31). 3-(3-Bromopropyl) pyridine (100 mg, 0.5 mmol) was treated with 1.35 g (1.5 mmol) tris(tetra-n-butylammonium) hydrogen diphosphate in CH_3CN (3 mL). Flash chromatography on a cellulose column (4.5:2.5:3.0 (v/v/v) 2-propanol/CH_3CN/50 mM NH_4HCO_3) yielded 57 mg (33 %) of a white solid. ^{1}H NMR (400 MHz, D_2O): δ 1.75-1.89 (m, 2H), 2.65-2.73 (m, 2H), 3.84 (t, J = 6.0 Hz, 2H), 7.34-7.36 (m, 1H), 7.74 (d, J = 8.0 Hz, 1H), 8.26 (s, 1H), 8.35 (s, 1H). ^{31}P NMR (162 MHz, D_2O): δ -9.49 (d, J = 20.7 Hz), -6.87 (d, J = 21.9 Hz).

5.6 Tables and Figures

Table 5.1. g-values reported for Fe_4S_4 proteins and those observed in this study.

	g_1	g_2	g_3
E. coli LytB Fe_4S_4(11)	2.038	1.927	1.889
E. coli GcpE Fe_4S_4(53)	2.04	1.90	–
Bovine heart aconitase (active form, Fe_4S_4)(54)	2.06	1.93	1.86
Bovine heart aconitase (inactive form, Fe_3S_4)(54)	–	2.016	–
E. coli GcpE $Fe_3S_{4(55)}$	2.032	2.003	–
Unliganded *A. aeolicus* LytB (Fig. 1a; only g values in the g=2 region are listed)	2.04	1.94; 1.92	1.87
Unliganded *E. coli* LytB (Fig. 1b). g value obtained by simulation.	2.043	1.915	1.901
A. aeolicus LytB with KCN (Fig. 1a)	2.08; 2.05	1.94; 1.93	–
A. aeolicus LytB with **12** (Fig. 1a). g value obtained by simulation.	2.087	2.012	1.992
P. furiosus ferredoxin (21)	2.12	1.86	1.79
P. furiosus ferredoxin with KCN (21)	2.09	1.95	1.92

Figure 5.1. 9 GHz EPR spectra of IspH at 15K reduced with $Na_2S_2O_4$. a) *A. aeolicus* IspH, no added ligands, microwave power = 1 mW. b) *E. coli* IspH, no added ligands, microwave power = 1 mW; c) *A. aeolicus* IspH + KCN, 10 equivalents, microwave power = 1 mW; d) *A. aeolicus* IspH + propiolic acid, 10 equivalents, microwave power = 0.2 mW; e) *A. aeolicus* IspH + but-3-ynyl diphosphate (**12**), 20 equivalents, microwave power = 0.05 mW. Spectral simulations are shown by the dashed lines (g-values are compiled in Table 5.1, together with g-values for Fe_3S_4 proteins).

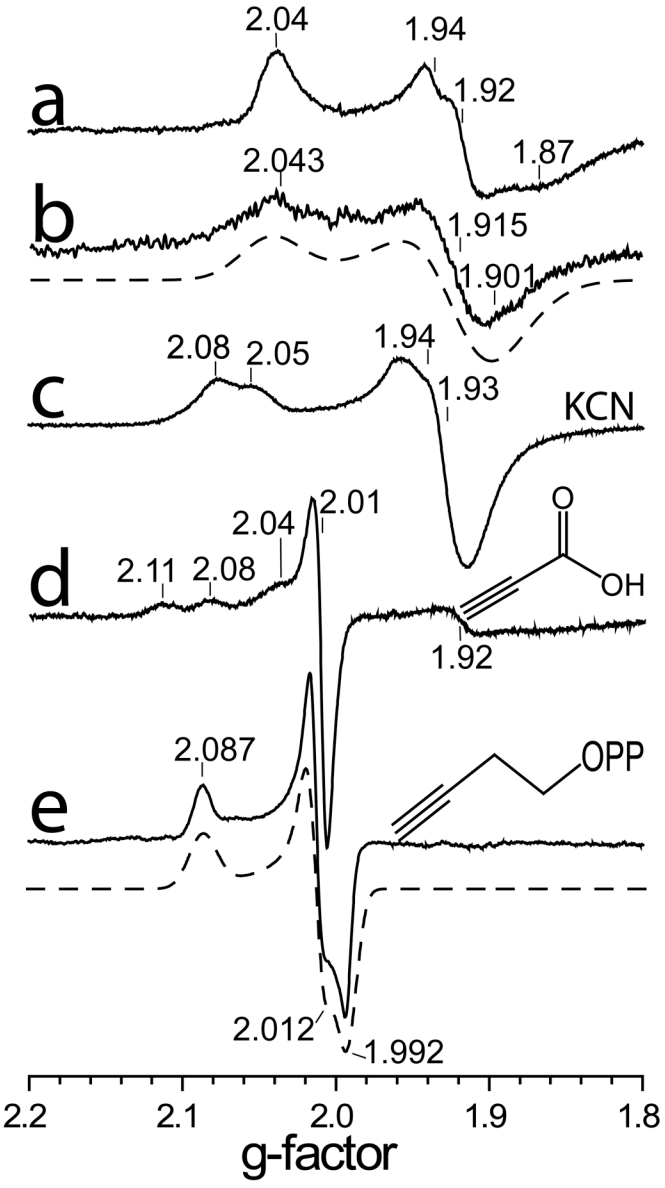


Figure 5.2. Wide scan X-band GHz EPR spectra of IspH at 15K, reduced with $Na_2S_2O_4$. a) *A. aeolicus* IspH, no added ligands, microwave power = 1 mW. b) *A. aeolicus* IspH + KCN, 10 equivalents, microwave power = 1 mW; c) *A. aeolicus* IspH + **12**, 20 equivalents, microwave power = 0.05 mW; d) *A. aeolicus* IspH + **26**, 20 equivalents, microwave power = 1 mW.

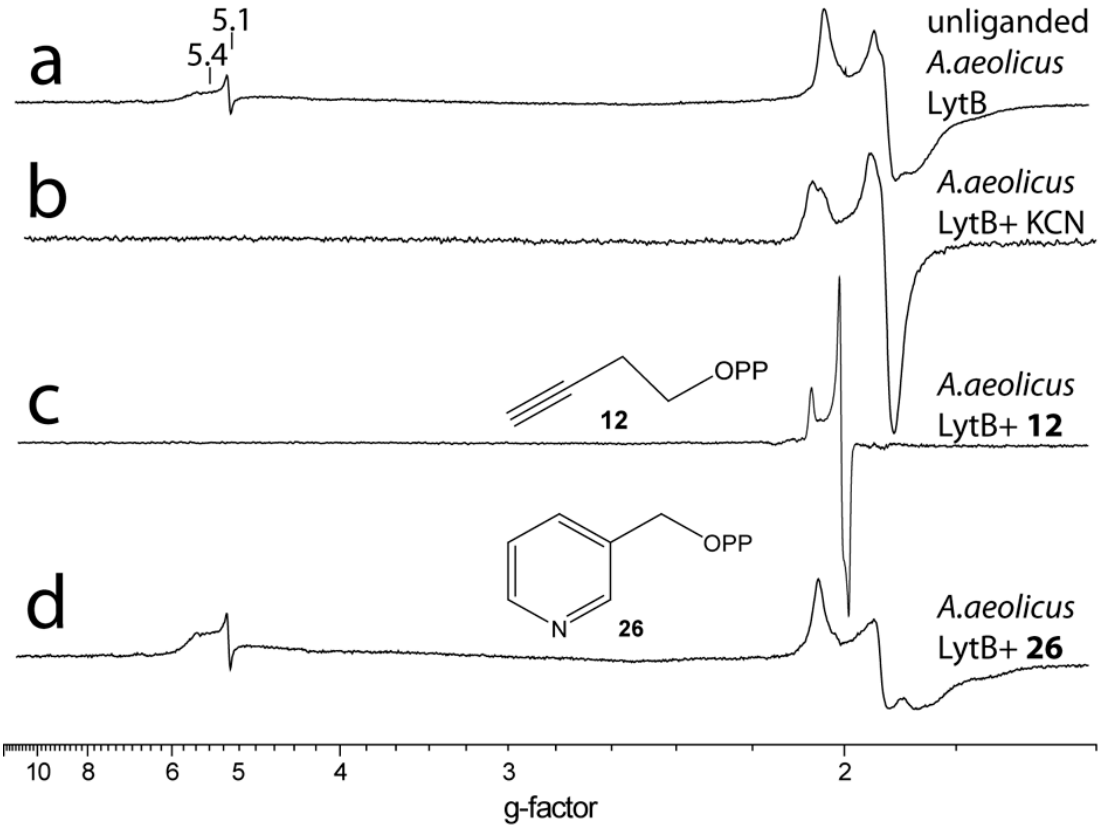


Figure 5.3. Structures of the 21 compounds investigated as *A. aeolicus* IspH inhibitors together with (below the structures) their IC_{50} values.

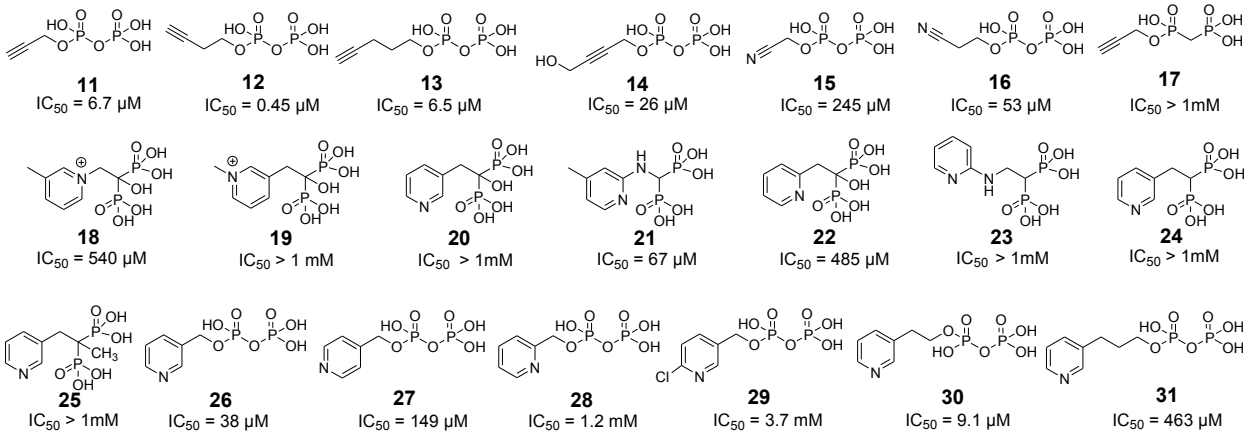


Figure 5.4. IspH inhibition by alkyne diphosphates. a) Dose-response curves for IspH inhibition by **12** (IC_{50} = 450 nM) and **13** (IC_{50} = 6.5 μM). b) Proposed docking model for **12** bound to the IspH active site: the diphosphate binds to the PPi site while the alkyne group forms a π (or π/σ) "metallacycle" complex with the unique 4th Fe in the Fe_4S_4 cluster, similar to that proposed for acetylene binding to model Fe_4S_4 clusters and allyl alcohol binding to a nitrogenase FeMo cofactor. c) αHOMO-1 for propargyl alcohol bound to a model Fe_4S_4 cluster, illustrating metal ligand interaction. d) as c) but βHOMO-1. The contour values are ± 0.02 au.

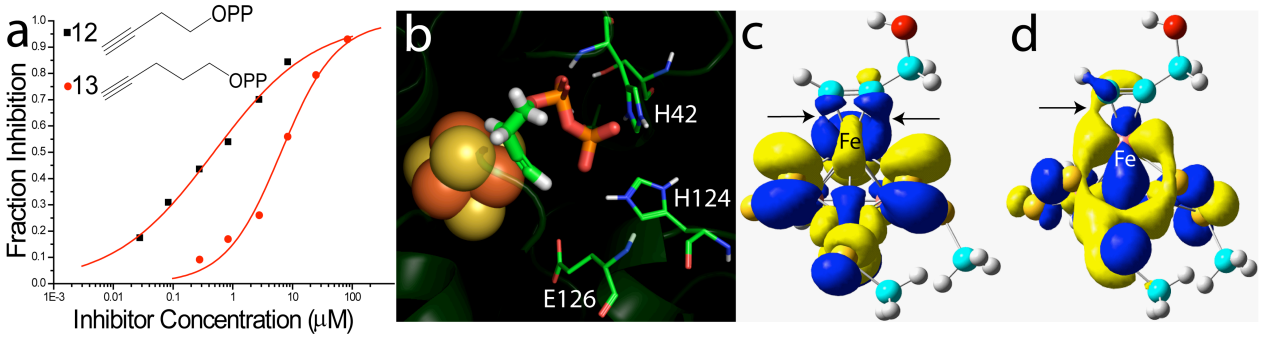


Figure 5.5. IspH inhibition by pyridinium diphosphates. a) Dose-response curves for IspH inhibition by **26**, **27**, **28** and **30**. The IC_{50} for the most potent inhibitor (**30**) is IC_{50} = 9.1 μM; K_i ~ 1.2 μM. b) Glide docking pose for **26** bound to the IspH active site: the diphosphate binds to the PPi site while the pyridinium sidechain interacts with E126. c) View of the *ortho*, *meta* and *para*-pyridinium diphosphates (**26**-**28**) docked to IspH: the most potent inhibitor has the shortest H^N…E126 CO_2^- distance; the worst inhibitor has the longest distance. d) Superimposed docking poses for **12** and **26** showing that they both bind in the same region (preventing HMBPP binding).

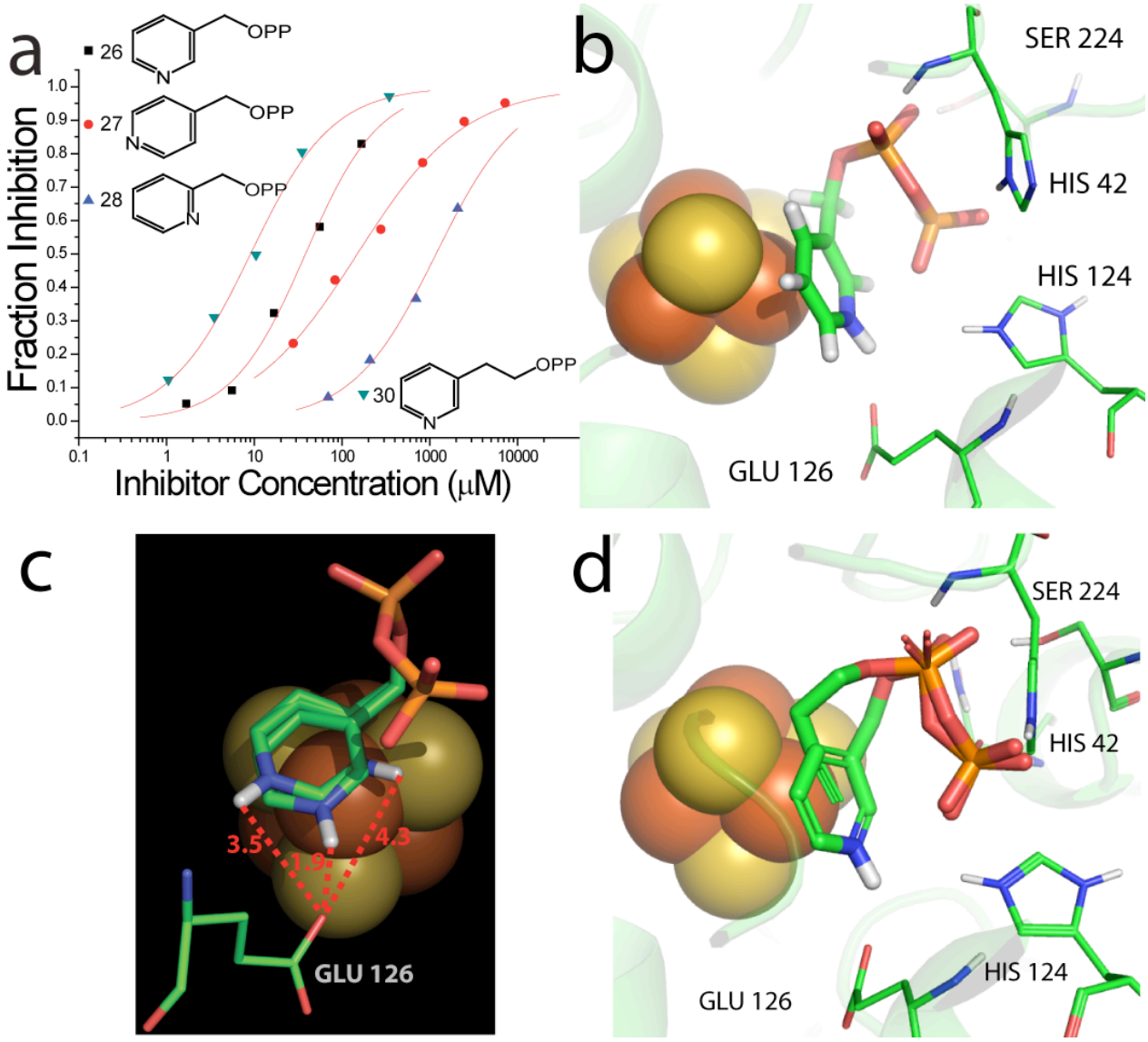


Figure 5.6. Mechanistic proposals for IspH catalysis and inhibition. a) π (or π/σ) metallacycle complex formation by alkyne inhibitors. b) IspH catalysis is which an η^2-alkenyl metallacycle is reduced and protonated then dehydrates to form an allyl complex, which is then cleaved to form the DMAPP and IPP products. c) IspH inhibition by pyridinium diphosphates which block HMBPP binding to the 4th Fe.

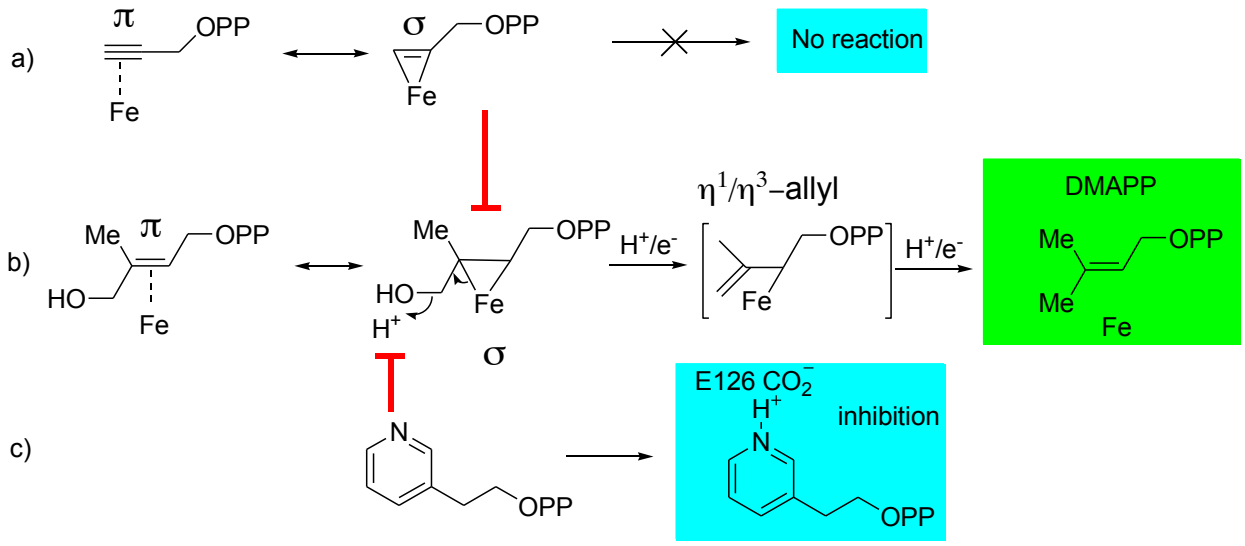
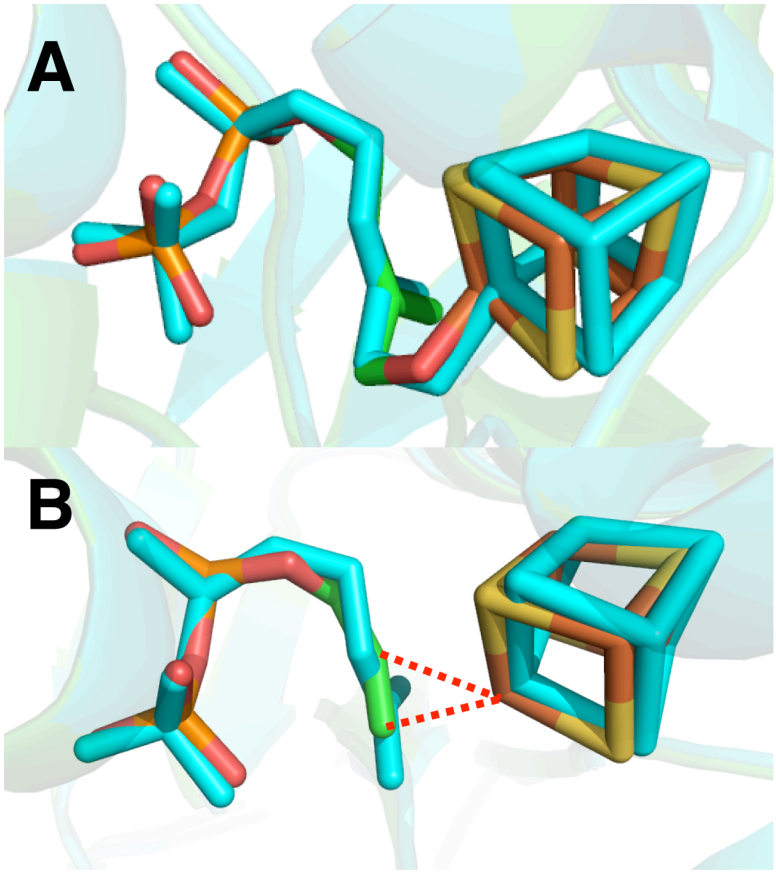


Figure 5.7. Comparisons between x-ray structures and docking /EPR/ENDOR deduced structures for **a**), HMBPP initial docking (PDB code 3KE8 and Ref. 25, Fig 1G). **b**) propargyl diphosphate (docking) superimposed on deoxy complex (PDB ID code 3KE9 and Ref. 25, Fig.



5.7 References

1. Rohmer M (2008) From Molecular Fossils of Bacterial Hopanoids to the Formation of Isoprene Units: Discovery and Elucidation of the Methylerythritol Phosphate Pathway. (Translated from English) *Lipids* 43(12):1095-1107 (in English).
2. Wiesner J & Jomaa H (2007) Isoprenoid biosynthesis of the apicoplast as drug target. (Translated from English) *Current Drug Targets* 8(1):3-13 (in English).
3. Williams SL & McCammon AJ (2009) Conformational dynamics of the flexible catalytic loop in *Mycobacterium tuberculosis* 1-deoxy-D-xylulose 5-phosphate reductoisomerase. (Translated from eng) *Chemical biology & drug design* 73(1):26-38 (in eng).
4. de Ruyck J & Wouters J (2008) Structure-based drug design targeting biosynthesis of isoprenoids: a crystallographic state of the art of the involved enzymes. (Translated from eng) *Current protein & peptide science* 9(2):117-137 (in eng).
5. Jomaa H, *et al.* (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. (Translated from English) *Science* 285(5433):1573-1576 (in English).
6. Borrmann S, *et al.* (2006) Fosmidomycin plus clindamycin for treatment of pediatric patients aged 1 to 14 years with *Plasmodium falciparum* malaria. (Translated from English) *Antimicrobial Agents and Chemotherapy* 50(8):2713-2718 (in English).
7. Rekittke I, *et al.* (2008) Structure of (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase, the terminal enzyme of the non-mevalonate pathway. (Translated from eng) *J Am Chem Soc* 130(51):17206-17207 (in eng).

8. Grawert T*, et al.* (2009) Structure of active IspH enzyme from *Escherichia coli* provides mechanistic insights into substrate reduction. (Translated from eng) *Angewandte Chemie (International ed* 48(31):5756-5759 (in eng).

9. Xiao Y, Chu L, Sanakis Y, & Liu P (2009) Revisiting the IspH catalytic system in the deoxyxylulose phosphate pathway: achieving high activity. (Translated from eng) *J Am Chem Soc* 131(29):9931-9933 (in eng).

10. Seemann M*, et al.* (2009) Isoprenoid Biosynthesis via the MEP Pathway: *In Vivo* Mossbauer Spectroscopy Identifies a [4Fe-4S](2+) Center with Unusual Coordination Sphere in the LytB Protein. (Translated from Eng) *J Am Chem Soc* 131(37):13184–13185 (in Eng).

11. Wolff M*, et al.* (2003) Isoprenoid biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB/IspH) from *Escherichia coli* is a [4Fe-4S] protein. (Translated from eng) *FEBS Lett* 541(1-3):115-120 (in eng).

12. Mulliez E*, et al.* (1999) Iron-sulfur interconversions in the anaerobic ribonucleotide reductase from *Escherichia coli*. (Translated from English) *J. Biol. Inorg. Chem.* 4(5):614-620 (in English).

13. Van Hoof S*, et al.* (2008) Synthesis of analogues of (E)-1-Hydroxy-2-methylbut-2-enyl 4-diphosphate, an isoprenoid precursor and human gamma delta T cell activator. (Translated from English) *J Org Chem* 73(4):1365-1370 (in English).

14. McMillan RS, Renaud J, Reynolds JG, & Holm RH (1979) Biologically related iron-sulfur clusters as reaction centers. Reduction of acetylene to ethylene in systems based on

[Fe4S4(SR)4]3. (Translated from eng) *Journal of inorganic biochemistry* 11(3):213-227 (in eng).

15. Itoh T, Nagano T, & Hirobe M (1980) [Fe4s4(Sr)4]-2- Catalytic Reduction of Diphenylacetylene to Cis-Stilbene in the Presence of Nabh4. (Translated from English) *Tetrahedron Letters* 21(14):1343-1346 (in English).

16. Tanaka K, Nakamoto M, Tsunomori M, & Tanaka T (1987) Raman-Spectra of the Adducts of Reduced Species of [Fe4s4(Sph)4]2- and [Mo2fe6s8(Sph)9]3- with Acetylene. (Translated from English) *Chem Lett* (4):613-616 (in English).

17. Yu Y, Smith JM, Flaschenriem CJ, & Holland PL (2006) Binding affinity of alkynes and alkenes to low-coordinate iron. (Translated from eng) *Inorg Chem* 45(15):5742-5751 (in eng).

18. Pelmenschikov V, Case DA, & Noodleman L (2008) Ligand-bound S=1/2 FeMo-cofactor of nitrogenase: Hyperfine interaction analysis and implication for the central ligand X identity. *Inorg. Chem.* 47(14):6162-6172.

19. Schunn RA, Fritchie CJ, Jr., & Prewitt CT (1966) Syntheses of some cyclopentadienyl transition metal sulfides and the crystal structure of $(C_5H_5FeS)_4$. *Inorg Chem* 5(5):892 - 899.

20. Crabtree RH (2005) *Chapter 3 and Chapter 5, The Organometallic Chemistry of the Transition Metals (Fourth Edition)* (John Wiley & Sons, Inc., Hoboken, New Jersey).

21. Telser J*, et al.* (1995) Cyanide Binding to the Novel 4Fe Ferredoxin from *Pyrococcus furiosus*: Investigation by EPR and ENDOR Spectroscopy. *J. Am. Chem. Soc.* 117:5133-5140.

22. Sanders JM*, et al.* (2005) Pyridinium-1-yl bisphosphonates are potent inhibitors of farnesyl diphosphate synthase and bone resorption. *J Med Chem* 48(8):2957-2963.

23. Martin MB, Arnold W, Heath HT, 3rd, Urbina JA, & Oldfield E (1999) Nitrogen-containing bisphosphonates as carbocation transition state analogs for isoprenoid biosynthesis. *Biochem. Biophys. Res. Commun.* 263(3):754-758.

24. Altincicek B*, et al.* (2002) LytB protein catalyzes the terminal step of the 2-C-methyl-D-erythritol-4-phosphate pathway of isoprenoid biosynthesis. (Translated from eng) *FEBS Lett* 532(3):437-440 (in eng).

25. Anonymous (2007) *Glide 4.5* (Schrodinger, LLC, New York).

26. Wang W*, et al.* (2010) Bioorganometallic Mechanism of Action, and Inhibition, of IspH. *Proc Natl Acad Sci USA* In press.

27. Grawert T*, et al.* (2010) Probing the reaction mechanism of IspH protein by x-ray structure analysis. *Proc Natl Acad Sci USA* 107:1077-1081.

28. Lee HI*, et al.* (2004) An organometallic intermediate during alkyne reduction by nitrogenase. *J. Amer. Chem. Soc.* 126(31):9563-9569.

29. Cheng Y & Prusoff WH (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. (Translated from eng) *Biochemical pharmacology* 22(23):3099-3108 (in eng).

30. Stoll S & Schweiger A (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. (Translated from eng) *J Magn Reson* 178(1):42-55 (in eng).

31. Anonymous (2007) *Maestro 8.0* (Schrodinger, LLC, New York).

32. Anonymous (2007) *MacroModel 9.5* (Schrodinger, LLC, New York).

33. Halgren TA (1996) Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. (Translated from English) *J Comput Chem* 17(5-6):490-519 (in English).

34. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38(6):3098-3100.

35. Perdew JP, Burke K, & Wang Y (1996) Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys Rev B Condens Matter* 54(23):16533-16539.

36. Wachters A (1970) *J. Chem. Phys.* 52:1033-1036.

37. Frisch M (2004) *Gaussian 03, Revision D.01* (Gaussian, Inc., Wallingford CT).

38. Zhang Y, Gossman W, & Oldfield E (2003) A density functional theory investigation of Fe-N-O bonding in heme proteins and model systems. (Translated from English) *J Am Chem Soc* 125(52):16387-16396 (in English).

39. Zhang Y, Mao J, Godbout N, & Oldfield E (2002) Mossbauer quadrupole splittings and electronic structure in heme proteins and model systems: a density functional theory investigation. (Translated from eng) *J Am Chem Soc* 124(46):13921-13930 (in eng).

40. Zhang Y, Mao JH, & Oldfield E (2002) Fe-57 Mossbauer isomer shifts of heme protein model systems: Electronic structure calculations. (Translated from English) *J Am Chem Soc* 124(26):7829-7839 (in English).

41. Zhang Y & Oldfield E (2003) Fe-57 Mossbauer quadrupole splittings and isomer shifts in spin-crossover complexes: A density functional theory investigation. (Translated from English) *J Phys Chem A* 107(20):4147-4150 (in English).

42. Zhang Y & Oldfield E (2003) An investigation of the unusual Fe-57 Mossbauer quadrupole splittings and isomer shifts in 2 and 3-coordinate Fe(II) complexes. (Translated from English) *J Phys Chem B* 107(29):7180-7188 (in English).

43. Zhang Y, Sun H, & Oldfield E (2005) Solid-state NMR fermi contact and dipolar shifts in organometallic complexes and metalloporphyrins. (Translated from eng) *J Am Chem Soc* 127(11):3652-3653 (in eng).

44. Davisson VJ, Woodside AB, & Poulter CD (1985) Synthesis of allylic and homoallylic isoprenoid pyrophosphates. (Translated from eng) *Methods in enzymology* 110:130-144 (in eng).

45. Jackson W, Perlmutter P, & Smallridge A (1988) The Stereochemistry of Organometallic Compounds. XXXII. Hydrocyanation of Derivatives of Amino Alkynes *Aust J Chem* 41(8):1201-1208.

46. Franceschin M*, et al.* (2007) Specific interactions with intra- and intermolecular G-quadruplex DNA structures by hydrosoluble coronene derivatives: A new class of telomerase inhibitors. *Bioorg Med Chem* 15(4):1848-1858.

47. Martin MB*, et al.* (2001) Bisphosphonates inhibit the growth of *Trypanosoma brucei, Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, and *Plasmodium falciparum*: a potential route to chemotherapy. *J Med Chem* 44(6):909-916.

48. Martin MB*, et al.* (2002) Activity of bisphosphonates against *Trypanosoma brucei* rhodesiense. *J Med Chem* 45(14):2904-2914.

49. Ghosh S*, et al.* (2004) Effects of bisphosphonates on the growth of *Entamoeba histolytica* and Plasmodium species *in vitro* and *in vivo*. *J Med Chem* 47(1):175-187.

50. Sanders JM*, et al.* (2004) Quantitative structure-activity relationships for gamma delta T cell activation by bisphosphonates. (Translated from English) *Journal of Medicinal Chemistry* 47(2):375-384 (in English).

51. Szabo CM, Martin MB, & Oldfield E (2002) An investigation of bone resorption and Dictyostelium discoideum growth inhibition by bisphosphonate drugs. (Translated from English) *Journal of Medicinal Chemistry* 45(14):2894-2903 (in English).

52. Davisson VJ*, et al.* (1986) Phosphorylation of Isoprenoid Alcohols. (Translated from English) *Journal of Organic Chemistry* 51(25):4768-4779 (in English).

53. Xiao Y, Zahariou G, Sanakis Y, & Liu P (2009) IspG enzyme activity in the deoxyxylulose phosphate pathway: roles of the iron-sulfur cluster. *Biochemistry* 48(44):10483-10485.

54. Beinert H, Kennedy MC, & Stout CD (1996) Aconitase as Ironminus signSulfur Protein, Enzyme, and Iron-Regulatory Protein. *Chem Rev* 96(7):2335-2374.

55. Grawert T*, et al.* (2004) IspH protein of *Escherichia coli*: studies on iron-sulfur cluster implementation and catalysis. (Translated from eng) *J Am Chem Soc* 126(40):12847-12855 (in eng).